



US009068201B2

(12) **United States Patent**  
**Hu et al.**

(10) **Patent No.:** US 9,068,201 B2  
(45) **Date of Patent:** \*Jun. 30, 2015

(54) **METHODS AND COMPOSITIONS FOR PRODUCING FATTY ALCOHOLS**(71) Applicant: **Reg Life Sciences, LLC**, Ames, IA (US)(72) Inventors: **Zhihao Hu**, South San Francisco, CA (US); **Vikranth Arlagadda**, South San Francisco, CA (US)(73) Assignee: **REG Life Sciences, LLC.**, South San Francisco, CA (US)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **13/647,185**(22) Filed: **Oct. 8, 2012**(65) **Prior Publication Data**

US 2013/0115665 A1 May 9, 2013

**Related U.S. Application Data**

(63) Continuation-in-part of application No. 12/575,430, filed on Oct. 7, 2009.

(60) Provisional application No. 61/109,131, filed on Oct. 28, 2008.

(51) **Int. Cl.**

<b>C12N 9/02</b>	(2006.01)
<b>A61K 31/045</b>	(2006.01)
<b>A61K 9/42</b>	(2006.01)
<b>C12P 7/04</b>	(2006.01)
<b>C12N 1/20</b>	(2006.01)
<b>A61K 38/00</b>	(2006.01)
<b>C12R 1/01</b>	(2006.01)

(52) **U.S. Cl.**

CPC ... <b>C12P 7/04</b> (2013.01); <b>C12N 1/20</b> (2013.01);	
<b>C12N 9/0004</b> (2013.01); <b>C12N 2501/71</b> (2013.01);	
<b>C12N 2502/99</b> (2013.01); <b>A61K 31/045</b> (2013.01);	
<b>C12R 1/01</b> (2013.01)	

(58) **Field of Classification Search**

CPC .....	<b>C12N 2501/71</b> ; <b>C12N 2009/02</b> ; <b>C12N 9/0004</b> ; <b>C12N 2502/99</b>
-----------	--

See application file for complete search history.

(56) **References Cited**

## U.S. PATENT DOCUMENTS

5,000,000 A	3/1991	Ingram et al.
5,028,539 A	7/1991	Ingram et al.
5,424,202 A	6/1995	Ingram et al.
5,482,846 A	1/1996	Ingram et al.
5,602,030 A	2/1997	Ingram et al.
5,807,893 A	9/1998	Voelker et al.
6,583,266 B1 *	6/2003	Smith et al. .... 530/350
7,056,714 B2	6/2006	Rosazza et al.
7,118,896 B2	10/2006	Kalscheuer et al.
7,169,588 B2	1/2007	Burch et al.
7,425,433 B2	9/2008	Rosazza et al.

7,491,854 B2	2/2009	Binder
7,608,700 B2	10/2009	Klaenhammer et al.
8,110,670 B2 *	2/2012	Hu et al. .... 536/23.2
8,323,924 B2 *	12/2012	Schirmer et al. .... 435/41
2003/0064328 A1 *	4/2003	Friedel .... 430/350
2003/0097686 A1	5/2003	Knauf et al.
2003/0129601 A1	7/2003	Cole
2003/0233675 A1	12/2003	Cao et al.
2004/0180400 A1	9/2004	Rosazza et al.
2004/0197896 A1	10/2004	Cole
2005/0019863 A1 *	1/2005	Sarmientos et al. .... 435/69.1
2005/0250135 A1	11/2005	Klaenhammer et al.
2007/0281345 A1	12/2007	Binder
2008/0221310 A1	9/2008	O'Sullivan et al.
2008/0295388 A1	12/2008	Bazzani et al.
2009/0084025 A1	4/2009	Bhatia et al.
2009/0136469 A1 *	5/2009	Senin et al. .... 424/93.51
2009/0275097 A1	11/2009	Sun et al.
2010/0105955 A1	4/2010	Alibhai et al.
2010/0105963 A1	4/2010	Hu et al.
2010/0154293 A1	6/2010	Hom et al.
2010/0221798 A1	9/2010	Schirmer et al.
2010/0242345 A1	9/2010	Keasling et al.
2010/0249470 A1	9/2010	Schirmer et al.
2010/0251601 A1	10/2010	Hu et al.
2010/0274033 A1 *	10/2010	Sanchez-Riera et al. .... 554/1
2011/0097769 A1 *	4/2011	Del Cardayre et al. .... 435/134
2012/0040426 A1	2/2012	Sun et al.
2013/0084608 A1 *	4/2013	Szabo et al. .... 435/99

## FOREIGN PATENT DOCUMENTS

WO	WO 2004081226 A2 *	9/2004
WO	WO 2007/003736 A1	1/2007
WO	WO 2007/136762 A2	11/2007
WO	WO 2008/058788	5/2008
WO	WO 2008/119082 A2	10/2008
WO	WO 2009/042950 A1	4/2009

(Continued)

## OTHER PUBLICATIONS

SEQ Align 22 (2013) Attached sequence alignment for SEQ ID No. 2 from the search results, pp. 1-3.\*

Abbadì et al., "Knockout of the regulatory site of 3-ketoacyl-ACP synthase III enhances short-and medium-chain acyl-ACP synthesis", *Plant Journal*, 24(1): 1-9 (2000).Abdel-Hamid et al., "Coordinate Expression of the Acetyl Coenzyme A Carboxylase Genes, *accB* and *accC*, Is Necessary for Normal Regulation of Biotin Synthesis in *Escherichia coli*", *J. Bacteriol.*, 189:369-376 (2007).

(Continued)

Primary Examiner — Manjunath Rao

Assistant Examiner — Samuel Liu

(74) Attorney, Agent, or Firm — Brigitte A. Hajos; REG Life Sciences LLC

(57) **ABSTRACT**

Methods and compositions, including nucleotide sequences, amino acid sequences, and host cells, for producing fatty alcohols are described.

(56)

**References Cited****FOREIGN PATENT DOCUMENTS**

WO	WO 2009/140695	11/2009
WO	WO 2009/140696	11/2009
WO	WO 2010/042664	4/2010
WO	WO 2010/062480 A2	6/2010
WO	WO 2011/062987 A2	5/2011
WO	WO 2010/075483 A1	7/2011

**OTHER PUBLICATIONS**

- Abdel-Hamid et al., "Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*," *Microbiol.* 147(6):1483-98 (2001).
- Alper, et al., "Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?", *NRM* 7: 715-723 (2009).
- Atsumi et al., "Metabolic engineering for advanced biofuels production from *Escherichia coli*," *Current Opin.Biotech.*, 19:414-419 (2008).
- Atsumi et al., "Metabolic engineering of *Escherichia coli* for 1-butanol production", *Metabolic Engineering* 10:305-311 (2008).
- Atsumi et al., "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels", *Nature*, 451: 86-89 (2008).
- Barnes, Jr. et al., "Studies on the Mechanism of Fatty Acid Synthesis. XIX. Preparation and General Properties of Palmitoyl Thioesterase", *J. Biol. Chem.*, 243(11):2955-2962 (1968).
- Bergler et al., "Protein EnvM is the NADH-dependent Enoyl-ACP Reductase (FabI) of *Escherichia coli*," *J. Biol. Chem.*, 269(8): 5943-5946 (1994).
- Bergler et al., "The enoyl-[acyl-carrier-protein] reductase (FabI) of *Escherichia coli*, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA", *Eur. J. Biochem.* 242: 689-694 (1996).
- Berrios-Rivera et al., "The Effect of Increasing NADH Availability on the Redistribution of Metabolic Fluxes in *Escherichia coli* Chemostat Cultures", *Metabolic Engineering* 4: 230-237 (2002).
- Birge et al., "Acyl Carrier Protein. XVI. Intermediate Reactions of Unsaturated Fatty Acid Synthesis in *Escherichia coli* and Studies of fab B Mutants", *J.Biol.Chem.* 247(16): 4921-4929 (1972).
- Black et al., "Cloning, Sequencing, and Expression of the fadD Gene of *Escherichia coli* Encoding Acyl Coenzyme A Synthetase," *J. Biol. Chem.* 267(35): 25513-25520 (1992).
- Black, P., "Primary Sequence of the *Escherichia coli* fadL Gene Encoding an Outer Membrane Protein Required for Long-Chain Fatty Acid Transport", *J. Bacteriology* 173(2): 435-442 (1991).
- Black et al., "Mutational Analysis of a Fatty Acyl-Coenzyme A Synthetase Signature Motif Identifies Seven Amino Acid Residues That Modulate Fatty Acid Substrate Specificity", *J. Biol. Chem.* 272(8) 4896-4903 (1997).
- Black et al., "Long-Chain Acyl-CoA—Dependent Regulation of Gene Expression in Bacteria, Yeast and Mammals", *J. Nutrition*, 305S-309S (2000).
- Blanchard et al., "Overexpression and Kinetic Characterization of the Carboxyltransferase Component of Acetyl-CoA Carboxylase", *J.Biol.Chem.* 273(30): 19140-19145 (1998).
- Bonamore et al., "The desaturase from *Bacillus subtilis*, a promising tool for the selective olefination of phospholipids", *J.Biotechnology* 121: 49-53 (2006).
- Bond-Watts et al., "Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways", *Nature Chem Bio* 537: 1-6 (Suppl. S1-S28) (2011).
- Bonner et al., "Purification and Properties of Fatty Acyl Thioesterase I from *Escherichia coli*", *J.Biol.Chem.* 247(10): 3123-3133 (1972).
- Boonstra et al., "The udhA Gene of *Escherichia coli* Encodes a Soluble Pyridine Nucleotide Transhydrogenase", *J. Bacteriol.* 181(3): 1030-1034 (1999).
- Boulanger et al., "Purification and Structural and Functional Characterization of FhuA, a Transporter of the *Escherichia coli* Outer Membrane," *Biochemistry*, 35(45): 14216-14224 (1996).
- Bredwell et al., "Reactor Design Issues for Synthesis-Gas Fermentations", *Biotechnol. Prog.* 15: 834-844 (1999).
- Broun et al., "Catalytic Plasticity of Fatty Acid Modification Enzymes Underlying Chemical Diversity of Plant Lipids", *Science* 282: 1315-1317 (1998).
- Bunch et al., "The IdhA gene encoding the fermentative lactate dehydrogenase of *Escherichia coli*," *Microbiol.* 143(1):187-95 (1997).
- Cahoon et al., "Modification of the Fatty Acid Composition of *Escherichia coli* by Coexpression of a Plant Acyl-Acyl Carrier Protein Desaturase and Ferredoxin", *J.Bacteriol.* 178(3): 936-936 (1996).
- Cahoon et al., "Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position", *Proc. Natl. Acad. Sci.* 94: 4872-4877 (1997).
- Cahoon et al., "A Determinant of Substrate Specificity Predicted from the Acyl-Acyl Carrier Protein Desaturase of Developing Cat's Claw Seed", *Plant Physiol* 117: 593-598 (1998).
- Campbell et al., "The Enigmatic *Escherichia coli* neu Gene is yaffH" *J. Bacteriol.*, 184(13): 3759-3764 (2002).
- Campbell et al., "A New *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic β-oxidation pathway," *Mol. Microbiol.*, 47(3): 793-805 (2003).
- Campbell et al., "*Escherichia coli* FadR Positively Regulates Transcription of the fabB Fatty Acid Biosynthetic Gene", *J.Bacteriol.* 183(20): 5982-5990 (2001).
- Caviglia et al., "Rat Long Chain Acyl-CoA Synthetase 5, but Not 1, 2, 3, or 4, Complements *Escherichia coli* fadD," *J. Biol. Chem.*, 279(12): 11163-11169 (2004).
- Chang et al., "Genetic and Biochemical Analyses of *Escherichia coli* Strains Having a Mutation in the Structural Gene (poxB) for Pyruvate Oxidase." *J. Bacteriol.*, 154(2): 756-62 (1983).
- Chen, et al., "Biosynthesis of Ansatrienin (mycotrienin) and naphthomycin, Identification and Analysis of Two Separate Biosynthetic Gene Clusters in *Streptomyces collinus* Tu 1892," *Eur. J. Biochem.* 261, 1999, pp. 98-107, 10 pages.
- Cheng, J. et al., "Mammalian Wax Biosynthesis," *J. Biol. Chem.* 279(36):37798-37807, 2004.
- Cho et al., "Defective Export of a Periplasmic Enzyme Disrupts Regulation of Fatty Acid Synthesis," *J. Biol. Chem.*, 270: 4216-4219 (1995).
- Cho et al., "*Escherichia coli* Thioesterase I, Molecular Cloning and Sequencing of the Structural Gene and Identification as a Periplasmic Enzyme", *J.Biol.Chem.* 268(13:5): 9238-9245 (1993).
- Cho et al., "Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA", *Microbiology* 152: 2207-2219 (2006).
- Choi et al., "β-Ketoacyl-acyl Carrier Protein Synthase III (FabH) Is a Determining Factor in Branched-Chain Fatty Acid Biosynthesis" *J. of Bacteriology* 182(2): 365-370 (2000).
- Collister et al., "Modification of the petroleum system concept: Origins of alkanes and isoprenoids in crude oils" *AAPG Bulletin* 88(5):587-611 (2004).
- Conway et al., "Cloning and Sequencing of the Alcohol Dehydrogenase II Gene from *Zymomonas mobilis*" *J. Bacteriol.* 169(6): 2591-2597 (1987).
- Cropp, et al., "Identification of a Cyclohexylcarbonyl CoA Biosynthetic Gene Cluster and Application in the Production of Doramectin," *Nature Biotechnology*, vol. 18, Sep. 2000, 4 pages.
- da Silva et al., "Comparison of the Genomes of Two *Xanthomonas* Pathogens with Differing Host Specificities" *Nature*, 417: 459-463 (2002).
- Database EMBL (Online), Jul. 1996, "Synechococcus, PCC7942 Ribosomal Protein S1 of 30S Ribosome (rpsl), ORF271, ORF231, ORF341, Carboxyltransferase alpha subunit (accA), ORF245, ORF227, and GTP cyclohydrolase I (folE) genes, complete cds, and ORF205 gene, partial cds.." XP002564232, 4 pages.
- Datsenko, et al., "One-step Inactivation of Chromosomal Genes in *Escherichia coli* K-12 using PCR Products," *Proc. Natl. Acad. Sci.* USA 97,2000, pp. 6640-6645, 6 pages.
- Davis et al., "Inhibition of *Escherichia coli* Acetyl Coenzyme A Carboxylase by Acyl-Acyl Carrier Protein" *J.Bacteriol.* 183(4): 1499-1503 (2001).

(56)

**References Cited****OTHER PUBLICATIONS**

- Dehesh et al., "KAS IV: A 3-ketoacyl-ACP synthase from *Cuphea* sp. Is a medium chain specific condensing enzyme", *The Plant Journal* 15(3):383-390 (1998).
- Dehesh et al., "Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*" *The Plant Journal* 9(2): 167-172 (1996).
- Delay et al., "In Vivo Functional Analyses of the Type II Acyl Carrier Proteins of Fatty Acid Biosynthesis", *J.Biol.Chem.* 282: 20319-20328 (2007).
- Demendoza et al., "Thermal Regulation of Membrane Fluidity in *Escherichia coli*. Effects of Overproduction of P-Ketoacylacyl Carrier Protein Synthase I", *J.Biol.Chem.* 258(4):2098-2101 (1983).
- Denoya, et al. "A Second Branched-Chain  $\alpha$ -Keto Acid Dehydrogenase Gene Cluster (bkdFPGH) from *Streptomyces avermitilis*: Its Relationship to Avermectin Biosynthesis and the Construction of a bkdf Mutant Suitable for the Production of Novel Antiparasitic Avermectins," *Journal of Bacteriology*, Jun. 1995, pp. 3504-3511, 8 pages.
- Deveaux et al., "Genetic and Biochemical Characterization of a Mutation (fatA) That Allows trans Unsaturated Fatty Acids to Replace the Essential cis Unsaturated Fatty Acids of *Escherichia coli*" *J.Bacteriol.* 171(3):1562-1568 (1989).
- Doan et al., "Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli*" *J. Plant Physiology* 166:787-796 (2009).
- Domergue et al., "Acyl Carriers Used as Substrates by the Desaturases and Elongases Involved in Very Long-chain Polyunsaturated Fatty Acids Biosynthesis Reconstituted in Yeast" *J.Biol.Chem.* 278(37):35115-35126 (2003).
- Domka et al., "YliH (BssR) and YceP (BssS) Regulate *Escherichia coli* K-12 Biofilm Formation by Influencing Cell Signaling" *Appl. and Environ. Microbiol.* 72(4):2449-2459 (2006).
- Doss, R.P., "Composition and Enzymatic Activity of the Extracellular Matrix Secreted by Germlings of *Botrytis cinerea*," *Appl. and Environ. Microbiol.*, 65(2): 404-408 (1999).
- Farewell et al., "Role of the *Escherichia coli* FadR Regulator in Stasis Survival and Growth Phase-Dependent Expression of the uspA, fad, and fab Genes", *J. Bacteriol.* 178(22): 6443-6450 (1996).
- Fehler et al., "Biosynthesis of Hydrocarbons in *Anabaena variabilis*. Incorporation of [ $methyl-^{14}C$ ]- and [ $methyl-^2H_3$ ] Methionine into 7- and 8-Methylheptadecanes", *Biochemistry* 9(2): 418-422 (1970).
- Feng et al., "A New Member of the *Escherichia coli* fad Regulon: Transcriptional Regulation of fadM (ybaW)", *J. Bacteriol.* 191(20): 6320-6328 (2009).
- Feng et al., "Overlapping Repressor Binding Sites Result in Additive Regulation of *Escherichia coli* FadH by FadR and ArcA" *J. of Bacteriology* 192(17):4289-4299 (2010).
- Feng et al., "Escherichia coli Unsaturated Fatty Acid Synthesis: Complex Transcription of the fabA Gene and in Vivo Identification of the Essential Reaction Catalyzed by FabB", *J.Biol.Chem.* 284(43): 29526-29535 (2009).
- Fischer et al., "Selection and optimization of microbial hosts for biofuels production" *Metabolic Engineering* 10:295-304 (2008).
- Fleischman et al., Accession No. YP\_889972/GI:11849671 (2006).
- Fozo et al., "The fabM Gene Product of *Streptococcus mutans* Is Responsible for the Synthesis of Monounsaturated Fatty Acids and Is Necessary for Survival at Low pH", *J. Bacteriol.* 186(13): 4152-4158 (2004).
- Ghislia et al., Acyl-CoA dehydrogenases—A mechanistic overview, *Eur. J. Biochem.* 271: 494-508 (2004).
- Hamilton-Kemp et al., "Production of the Long-Chain Alcohols Octanol, Decanol, and Dodecanol by *Escherichia coli*", *Current Microbiology* 51: 82-86 (2005).
- Han, et al., "A Novel Alternate Anaplerotic Pathway to the Glyoxylate Cycle in *Streptomyces*," *Journal of Bacteriology*, Aug. 1997, pp. 5157-5164, 8 pages.
- Han et al., "Biosynthesis of Alkanes in Nostroc Muscorum," *Journal of the American Chemical Society*, 91:18, Aug. 1969, pp. 5156-5159.
- He, et al., "Nocardia sp. Carboxylic Acid Reductase: Cloning, Expression, and Characterization of a New Aldehyde Oxidoreductase Family," *Applied and Environmental Microbiology*, Mar. 2004, pp. 1874-1881, 8 pages.
- Heath et al., "Regulation of Malonyl-CoA Metabolism by Acyl-Acyl Carrier Protein and  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthases in *Escherichia coli*", *J.Biol.Chem.* 270(26):15531-15538 (1995).
- Heath et al., "Regulation of Fatty Acid Elongation and Initiation by Acyl-Acyl Carrier Protein in *Escherichia coli*", *J.Biol.Chem.* vol. 271(4): 1833-1836 (1996).
- Heath et al., "Inhibition of  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase III (FabH) by Acyl-Acyl Carrier Protein in *Escherichia coli*", *J.Biol.Chem.* 271(18):10996-11000 (1996).
- Heath et al., "Roles of the FabA and FabZ  $\beta$ -Hydroxyacyl-Acyl Carrier Protein Dehydratases in *Escherichia coli* Fatty Acid Biosynthesis", *J.Biol.Chem.* 271(44): 27795-27801 (1996).
- Heath et al., "Lipid Biosynthesis as a Target for Antibacterial Agents," *Progress in Lipid Research* 40, 2001, pp. 467-497, 31 pages.
- Henry et al., "Escherichia coli Transcription Factor That Both Activates Fatty Acid Synthesis and Represses Fatty Acid Degradation", *J. Mol. Biol.* 222: 843-849 (1991).
- Hoffmeister, et al., "Mitochondrial trans-2-Enoyl-CoA Reductase of Wax Ester Fermentation from *Euglena gracilis* Defines a New Family of Enzymes Involved in Lipid Synthesis\*", *The Journal of Biological Chemistry*, vol. 280, No. 6, Issue of Feb. 2005, pp. 4329-4338, 10 pages.
- Hu et al., Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, *The Plant Journal* 54: 621-639 (2008).
- Hunt et al., "Characterization of an Acyl-CoA Thioesterase That Functions as a Major Regulator of Peroxisomal Lipid Metabolism" *J.Biol.Chem.* 277(2):1128-1138 (2002).
- Inui, et al., "Fatty Acid Synthesis in Mitochondria of *Euglena gracilis*," *Eur. J. Biochem.* 142, 1984, pp. 121-126, 6 pages.
- Jarboe, L.R. et al., "Development of Ethanogenic Bacteria," *Adv. Biochem. Engin./Biotechnol.* 108:237-261 (2007).
- Jayakumar et al., "Cloning and expression of the multifunctional human fatty acid synthase and its subdomains in *Escherichia coli*," *PNAS* 93: 14509-14514 (1996).
- Jiang et al., "Inhibition of Fatty Acid Synthesis in *Escherichia coli* in the Absence of Phospholipid Synthesis and Release of Inhibition by Thioesterase Action," *Journal of Bacteriology*, vol. 176, No. 10, May 1994, pp. 2814-2821.
- Johnson, et al., "Genetic Analysis of the Role of *Saccharomyces cerevisiae* Acyl-CoA Synthetase Genes in Regulating Protein N-Myristylation+", *The Journal of Biological Chemistry*, vol. 269, No. 27, Issue of Jul. 1994, pp. 18037-18046, 10 pages.
- Jones et al., "Palmitoyl-Acyl Carrier Protein (ACP) Thioesterase and the Evolutionary-Origin of Plant Acyl-ACP Thioesterases", *Plant Cell*, vol. 7:359-371 (1995).
- Juttner et al., "Environmental Factors Affecting the Formation of Mesityloide, Dimethylallylic Alcohol and Other Volatile Compounds Excreted by *Anabaena cylindrica*," *Journal of General Microbiology*, 1983, 129, pp. 407-412.
- Juttner et al., "The Reducing Capacities of Cyanobacteria for Aldehydes and Ketones," *Appl. Microbiol. Biotechnol.*, 1986, 25, pp. 52-54.
- Kalscheuer et al., "Microdiesel: *Escherichia coli* Engineered for Fuel Production," *Microbiology*, 2006, 152, pp. 2529-2536.
- Kalscheuer et al., "Neutral Lipid Biosynthesis in Engineered *Escherichia coli*: Jojoba Oil-Like Wax Esters and Fatty Acid Butyl Esters", *Appl. Environ. Microbiol.* 72(2): 1373-1379 (2006).
- Kalscheuer et al., "Analysis of Storage Lipid Accumulation in *Alcanivorax borkumensis*: Evidence for Alternative Triacylglycerol Biosynthesis Routes in Bacteria," *J.Bacteriology* 189(3): 918-923 (2007).
- Kalscheuer et al., "A Novel Bifunctional Wax Ester Synthase/Acyl-CoA:Diacylglycerol Acyltransferase Mediates Wax Ester and Triacylglycerol Biosynthesis in *Acinetobacter alcoaceticus* ADPI," *J. Biol. Chem.* 278:8075-8082, 2003.
- Kameda et al., "Further purification, characterization and salt activation of acyl-CoA synthetase from *Escherichia coli*", *Biochimica et Biophysica Acta* 840: 29-36(1985).

(56)

**References Cited**

## OTHER PUBLICATIONS

- Keasling et al., "Metabolic engineering delivers next-generation biofuels", *Nature Biotechnology* 26(3):298-299 (2008).
- Knoll, et al., "Biochemical Studies of Three *Saccharomyces cerevisiae* Acyl-CoA Synthetases, Faa1p, Faa2p, and Faa3p\*", The Journal of Biological Chemistry, vol. 269, No. 23, Issue of Jun. 1994, pp. 16348-16356, 9 pages.
- Knothe, G. "Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters," *Fuel Processing Technology* 86:1059-1070, 2005.
- Knudsen et al., "Transacylation as a chain-termination mechanism in fatty acid synthesis by mammalian fatty acid synthetase. Synthesis of medium-chain-length (C8-C12) acyl-CoA esters by goat mammary-gland fatty acid synthetase", *Biochem. J.* 202: 139-143 (1982).
- Koffas, M.A.G., "Expanding the repertoire of biofuel alternatives through metabolic pathway evolution", *PNAS* 106(4): 965-966 (2009).
- Kumari et al., "Regulation of Acetyl Coenzyme A Synthetase in *Escherichia coli*", *J. Bacteriol.* 182(15): 4173-4179 (2000).
- Ladygina et al., "A Review of Microbial Synthesis of Hydrocarbons," *Process Biochemistry*, 41, 2006, pp. 1001-1014.
- Lang et al., "Preparation and characterization of bio-diesels from various bio-oils", *Bioresource Tech.* 80: 53-62 (2001).
- Lee et al., "Enhanced preference for  $\pi$ -bond containing substrates is correlated to Pro110 in the substrate-binding tunnel of *Escherichia coli* thioesterase 1/protease 1/lysophospholipase L<sub>1</sub>" *Biochim. Et Biophys. Acta*, 1774: 959-967 (2007).
- Lee et al., "Metabolic engineering of microorganisms for biofuels production from bugs to synthetic biology to fuels", *Current Opinion in Biotechnology* 19: 556-563 (2008).
- Lennen et al., "A Process for Microbial Hydrocarbon Synthesis: Overproduction of Fatty Acids in *Escherichia coli* and Catalytic Conversion to Alkane", *Biotech.Bioengineering* 106(2):193-202 (2010).
- Leonard et al., "A *Cuphea*  $\beta$ -ketoacyl-ACP synthase shifts the synthesis of fatty acids towards shorter chains in *Arabidopsis* seeds expressing *Cuphea* FatB thioesterases", *Plant Journal* 13(5): 621-628 (1998).
- Li et al., "Alteration of the Fatty Acid Profile of *Streptomyces Coelicolor* by Replacement of the Initiation Enzyme 3-Ketoacyl Acyl Carrier Protein Synthase III (FabH)," *Journal of Bacteriology*, Jun. 2005, pp. 3795-3799, 5 pages.
- Li et al., "The Gene Encoding the Biotin Carboxylase Subunit of *Escherichia coli* Acetyl- CoA carboxylase", *J.Biol.Chem.* 267(2): 855-863 (1992).
- Li et al., "The carboxylic acid reduction pathway in *Nocardia*. Purification and characterization of the aldehyde reductase", *J. of Industrial Microbiology & Biotechnology* 25: 328-332 (2000).
- Li et al., "Conversion of Fatty Aldehydes to Alka(e)nes and Formate by a Cyanobacterial Aldehyde Decarbonylase: Cryptic Redox by an Unusual Dimetal Oxygenase", *J. Am. Chem. Soc.* 133: 6158-6161 (2011).
- Li et al., "Growth Rate Regulation of *Escherichia coli* Acetyl Coenzyme A Carboxylase, Which Catalyzes the First Committed Step of Lipid Biosynthesis", *J. Bacteriol.* 175(2): 332-340 (1993).
- Li et al., "Purification, Characterization, and Properties of an Aryl Aldehyde Oxidoreductase from *Nocardia* Sp. Strain NRRL 5646," *Journal of Bacteriology*, Jun. 1997, pp. 3482-3487, 6 pages.
- Link et al., "Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type *Escherichia coli*: Application to Open Reading Frame Characterization", *J.Bacteriol.* 179(20): 6228-6237 (1997).
- Liu, et al., "Production and secretion of fatty acids in genetically engineered cyanobacteria" *PNAS Early Edition*: 1-6 (2010).
- Lu et al., "Overproduction of free fatty acids in *E. coli*: Implications for biodiesel production", *Metabolic Engineering* 10: 333-339 (2008).
- Lykidis et al., "Genomic Prospecting for Microbial Biodiesel Production," NN, Jun. 2008, 41 pages.
- Mackey et al., "Detection of Rhythmic Bioluminescence from Luciferase Reporters in Cyanobacteria," *Methods in Molecular Biology*, Bol. 362, 2007, 16 pages.
- Magnuson et al., "Regulation of Fatty Acid Biosynthesis in *Escherichia coli*" *Microbiol.Reviews* 57(3): 522-542 (1993).
- Marr, et al., "Effect of Temperature on the Composition of Fatty Acids in *Escherichia coli*", *J.Bacteriol.* 84: 1260-1267 (1962).
- Marrakchi, et al., "A New Mechanism for Anaerobic Unsaturated Fatty Acid Formation in *Streptococcus pneumoniae*\*," *The Journal of Biological Chemistry*, Bol. 277, No. 47, Issue of Nov. 2002, pp. 44809-44816, 6 pages.
- Marrakchi, et al., "Mechanistic Diversity and Regulation of Type II Fatty Acid Synthesis," *Biochemical Society Transactions*, 2002, vol. 30, Part 6, pp. 1050-1055, 6 pages.
- Massengo-Tiasse et al., "Vibrio neumon FabV Defines a New Class of Enoyl-Acyl Carrier Protein Reductase", *J.Biol.Chem.* 283(3): 1308-1316 (2008).
- Mat-Jan et al., "Mutants of *Escherichia coli* Deficient in the Fermatative Lactate Dehydrogenase," *J. Bacteriol.* 171(1):342-8 (1989).
- Mayer et al., "Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach" *BMC Plant Biology* 7(1) (2007).
- McCue, L. et al., "hylogenetic footprinting of transcription factor binding sites in proteobacterial genomes," *Nucleic Acids Res.*, 29(3):774-82(2001).
- Metz et al., "Purification of a Jojoba Embryo Fatty Acyl-Coenzyme A Reductase and Expression of Its cDNA in High Erucic Acid Rapeseed", *Plant Physiol.* 122: 635-644 (2000).
- Metzgar et al., "Acinetobacter sp. ADP1: an ideal model organism for genetic analysis and genome engineering", *Nucleic Acid Res.* 32(19):5780-5790 (2004).
- Mohan et al., "An *Escherichia coli* Gene (FabZ) Encoding (3R)-Hydroxymyristoyl Acyl Carrier Protein Dehydrase. Relation to fabA and Suppression of Mutations in Lipid A Biosynthesis", *J.Biol.Chem.* 269(52): 32896-32903 (1994).
- Morgan-Kiss et al., "The *Escherichia coli* fadK (*ydiD*) Gene Encodes an Aerobically Regulated Short Chain Acyl-CoA Synthetase\*", *The Journal of Biological Chemistry*, vol. 279, No. 36, Sep. 2004, pp. 37324-37333.
- Morgan-Kiss et al., "The *Lactococcus lactis* FabF Fatty Acid Synthetic Enzyme can Functionally Replace both the FabB and FabF Proteins of *Escherichia coli* and the FabH Protein of *Lactococcus lactis*," *Arch. Microbiol.* 190: 427-437 (2008).
- Naccarato et al., "In Vivo and In Vitro Biosynthesis of Free Fatty Alcohols in *Escherichia coli* K-12," *Lipids*, vol. 9, No. 6, 1973, 10 pages.
- NCBI Reference, Putative Alcohol Dehydrogenase [Acinetobacter sp. ADP1], 2010, retrieved from <http://ncbi.nlm.nih.gov/protein/49532534>.
- NCBI Reference Sequence YP\_889972.1, Putative Long-Chain Fatty-Acid-CoA Ligase [Microbacterium Smegmatis Str. MC2 155], retrieved from <http://www.ncbi.nlm.nih.gov/protein/118469671>, 4 pages.
- Nunn et al., "Transport of long-chain fatty acids by *Escherichia coli*: Mapping and characterization of mutants in the fadL gene" *PNAS* 75(7): 3377-3381 (1978).
- Nunn et al., "Role for fadR in Unsaturated Fatty Acid Biosynthesis in *Escherichia coli*", *J.Bacteriol.* 154(2):554-560 (1983).
- Palaniappan et al., "Enhancement and Selective Production of Phoslactomycin B, a Protein Phosphatase Ila Inhibitor, through Identification and Engineering of the Corresponding Biosynthetic Gene Cluster\*", *The Journal of Biological Chemistry*, vol. 278, No. 37, Issue of Sep. 2003, pp. 35552-35557, 6 pages.
- Patton et al., "A Novel  $\Delta^3$ ,  $\Delta^2$ -Enoyl-CoA Isomerase Involved in the Biosynthesis of the Cyclohexanecarboxylic Acid-Derived Moiety of the Polyketide Ansatrienin A," *Biochemistry* 2000, 39, pp. 7595-7604, 10 pages.
- Peng et al., "Effect of fadR gene knockout on the metabolism of *Escherichia coli* based on analyses of protein expressions, enzyme activities and intracellular metabolite concentrations" *Enzyme and Microbial Tech.* 38: 512-520 (2006).

(56)

**References Cited**

## OTHER PUBLICATIONS

- Perez et al., "Escherichia coli YqhD Exhibits Aldehyde Reductase Activity and Protects from the Harmful Effect of Lipid Peroxidation-derived Aldehydes" *J.Biol.Chem.* 283(12): 7346-7353 (2008).
- Peterson & Ingram, "Anaerobic Respiration in Engineered *Escherichia coli* with an Internal Electron Acceptor to Produce Fuel Ethanol," *Ann. N.Y. Acad. Sci.* 1125:363-372 (2008).
- Phung et al., "Genes for Fatty Acids Biosynthesis in the Cyanobacterium *Synechococcus* sp. Strain PCC 7942," Jan. 1995, Abstracts of the General Meeting of the American Society of Microbiology, The Society, Washington, DC, p. 524, 1 page.
- Pillai et al., "Functional characterization of  $\beta$ -ketoacyl-ACP reductase (FabG) from *Plasmodium falciparum*" *Biochem. and Biophysical Research Comm.* 303: 387-392 (2003).
- Qiu et al., "Crystal structure and substrate specificity of the  $\beta$ -ketoacyl-acyl carrier protein synthase III (FabH) from *Staphylococcus aureus*", *Protein Science* 14: 2087-2094 (2005).
- Rafi et al., "Structure of Acyl Carrier Protein Bound to FabI, the FASII Enoyl Reductase from *Escherichia coli*" *J.Biol.Chem.* 281(51): 39285-39293 (2006).
- Rawlings et al., "The Gene Encoding *Escherichia coli* Acyl Carrier Protein Lies within a Cluster of Fatty Acid Biosynthetic Genes", *J.Biol.Chem.* 267(9):5751-5754 (1992).
- Rawlings et al., "Biosynthesis of fatty acids and related metabolites", *Natural Product Reports* 15: 275-308 (1998).
- Ray et al., "Activation of long chain fatty acids with acyl carrier protein: Demonstration of a new enzyme, acyl-acyl carrier protein synthetase, in *Escherichia coli*" *PNAS* 73(12):4374-4378 (1976).
- Rehm et al., "Heterologous expression of the acyl-acyl carrier protein thioesterase gene from the plant *Umbellularia californica* mediates polyhydroxyalkanoate biosynthesis in recombinant *Escherichia coli*", *Appl. Microbiol. and Biotech.* 55: 205-209 (2001).
- Reiser, et al., "Isolation of Mutants of *Acinetobacter calcoaceticus* Deficient in Wax Ester Synthesis of Complementation of One Mutation with a Gene Encoding a Fatty Acyl Coenzyme A Reductase," *Journal of Bacteriology*, May 1997, pp. 2969-2975, 7 pages.
- Rock et al., "Acyl-Acyl Carrier Protein Synthetase from *Escherichia coli*", *Meth. Enzymol.* 71: 163-168 (1981).
- Rock, et al., Increased Unsaturated Fatty Acid Production Associated with a Suppressor of the fabA6(Ts) Mutation in *Escherichia coli*, *Journal of Bacteriology*, Sep. 1996, vol. 178, No. 18, pp. 5382-5387, 6 pages.
- Rude et al., "Terminal Olefin (1-Alkene) Biosynthesis by a Novel P450 Fatty Acid Decarboxylase from *Jeotgalicoccus* Species", *Appl. Environ.Microbiol.* 77(5): 1718-1727 (2011).
- Rude et al., "New microbial fuels: a biotech perspective", *Current Opinion in Microbiology* 12: 274-281 (2009).
- Sabirova et al., "Mutation in a "tesB-Like" Hydroxyacyl-Coenzyme A-Specific Thioesterase Gene Causes Hyperproduction of Extracellular Polyhydroxyalkanoates by *Alcanivorax borkumensis* SK2", *J. Bacteriol.* 188(23): 8452-8459 (2006).
- Saito et al., "Crystal structure of enoyl-acyl carrier protein reductase (FabK) from *Streptococcus pneumoniae* reveals the binding mode of an inhibitor", *Protein Science* 17: 691-699 ((2008).
- Salas et al., "Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases", *Archives of Biochem. and Biophysics* 403: 25-34 (2002).
- Sanchez et al., "Effect of Overexpression of a Soluble Pyridine Nucleotide Transhydrogenase (UdhA) on the Production of Poly(3-hydroxybutyrate) in *Escherichia coli*", *Biotechnol.Prog.* 22: 420-425 (2006).
- Schirmer et al., "Microbial Biosynthesis of Alkanes", *Science* 329:559-562 (2010).
- Schujman et al., "A malonyl-CoA-dependent switch in the bacterial response to a dysfunction of lipid metabolism", *Molecular Microbiology*, 68(4): 987-996 (2008).
- Schweizer et al., "Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems", *Microbiol. Mol.Biol.Rev.* 68(3): 501-517 (2004).
- Shockley et al., "Arabidopsis Contains Nine Long-Chain Acyl-Coenzyme A Synthetase Genes that Participate in Fatty Acid and Glycerolipid Metabolism," *Plant Physiology*, Aug. 2002, vol. 129, pp. 1710-1722, 13 pages.
- Spencer, et al., "Thioesterases I and II of *Escherichia coli*," *The Journal of Biological Chemistry*, vol. 253, No. 17, Issue Sep. 1978, pp. 5922-5926, 5 pages.
- Stöveken et al., "The Wax Ester Synthase/Acyl Coenzyme A:Diacylglycerol Acyltransferase from *Acinetobacter* sp. Strain ADP1: Characterization of a Novel Type of Acyltransferase", *J. Bacteriology* 187(4)1369-1376 (2005).
- Subrahmanyam et al., "Overproduction of a Functional Fatty Acid Biosynthetic Enzyme Blocks Fatty Acid Synthesis in *Escherichia coli*" *J.Bacteriology* 180(17):4596-4602 (1998).
- Sulzenbacher et al., "Crystal Structure of *E. coli* Alcohol Dehydrogenase YqhD: Evidence of a Covalently Modified NADP Coenzyme" *J.Mol.Biol.* 342: 489-502 (2004).
- Ta et al., "Cloning, Sequencing, and Overexpression of [2Fe-2S] Ferredoxin Gene from *Escherichia coli*", *J.Biol.Chem.* 267(16):11120-11125 (1992).
- Thomason et al., "Identification of the *Escherichia coli* K-12 ybhE Gene as pgl, Encoding 6-Phosphogluconolactonase" *J.Bacteriol.* 186(24): 8248-8253 (2004).
- Thorpe et al., "Structure and mechanism of action of the Acyl-CoA dehydrogenases1", *FASEB J.* 9: 718-725 (1995).
- Tong et al., "Acetyl-Coenzyme A Carboxylases: Versatile Targets for Drug Discovery", *J. Cellular Biochem.* 99: 1476-1488 (2006).
- Toomey et al., "Studies on the Mechanism of Fatty Acid Synthesis XVI. Preparation and General Properties of Acyl-Malonyl Acyl Carrier Proteincondensing Enzyme From *Escherichia coli*", *J.Biol. Chem.* 241(5):1159-1165 (1996).
- Tsay et al., "Isolation and Characterization of the  $\beta$ -Ketoacyl-acyl Carrier Protein Synthase I11 Gene (fabH) from *Escherichia coli* K-12", *J.Biol.Chem.* 267(10): 6807-6814 (1992).
- Tucci et al., "A Novel Prokaryotic trans-2-enoyl-CoA reductase from the Spirochete *Treponema denticola*," *FEBS Letters* 581, 2007, pp. 1561-1566, 6 pages.
- UniProt accession No. Q325A2 "Subname: Full=Acyl-CoA thioesterase I" (2005).
- Vadali et al., "Cofactor engineering of intracellular CoA/acetyl-CoA and its effect on metabolic flux redistribution in *Escherichia coli*" *Metabolic Engineering* 6: 133-139 (2004).
- Van Den Berg et al., "The FadL family: unusual transporters for unusual substrates", *Curr.Opin.Struct.Biol.* 15: 401-407 (2005).
- Venkitasubramanian et al., "Reduction of Carboxylic Acids by *Nocardia* Aldehyde Oxidoreductase Requires a Phosphopantetheinylated Enzyme\*", *The Journal of Biological Chemistry*, vol. 282, No. 1, pp. 478-485, Jan. 2007, 8 pages.
- Voelker et al. "Alteration of the Specificity and Regulation of Fatty Acid Synthesis of *Escherichia coli* by Expression of a Plant Medium-Chain Acyl-Acyl Carrier Protein Thioesterase," *J. Bacteriol.*, 176(23): 7320-7327 (1994).
- Wang et al., "Functional Replacement of the FabA and FabB Proteins of *Escherichia coli* Fatty Acid Synthesis by *Enterococcus faecalis* FabZ and FabF Homologues", *J.Biol.Chem.* 279(33): 34489-34495 (2004).
- White et al., "Carboxylic acid reductase: a new tungsten enzyme catalyzes the reduction of non-activated carboxylic acids to aldehydes", *Eur. J. Biochem.* 184: 89-96 (1989).
- Xu et al., "The FadRzDNA Complex. Transcriptional Control of Fatty Acid Metabolism in *Escherichia coli*", *J.Biol.Chem.* 276(20): 17373-17379, 2001.
- Yomano, L.P. et al., "Re-Engineering *Escherichia coli* for ethanol production," *Biotechnol. Lett.* 30:2097-2103 (2008).
- Yoo et al., "Determination of the native form of FadD, the *Escherichia coli* fatty acyl-CoA synthetase, and characterization of limited proteolysis by outer membrane protease OmpT", *Biochem. J.* 360: 699-706 (2001).
- Yuan-Zheng et al., Metabolic Engineering of *Aeromonas hydrophila* for the Enhanced Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate), *Appl. Microbiol. Biotechnol.*, 2006, 69, pp. 537-532.

(56)

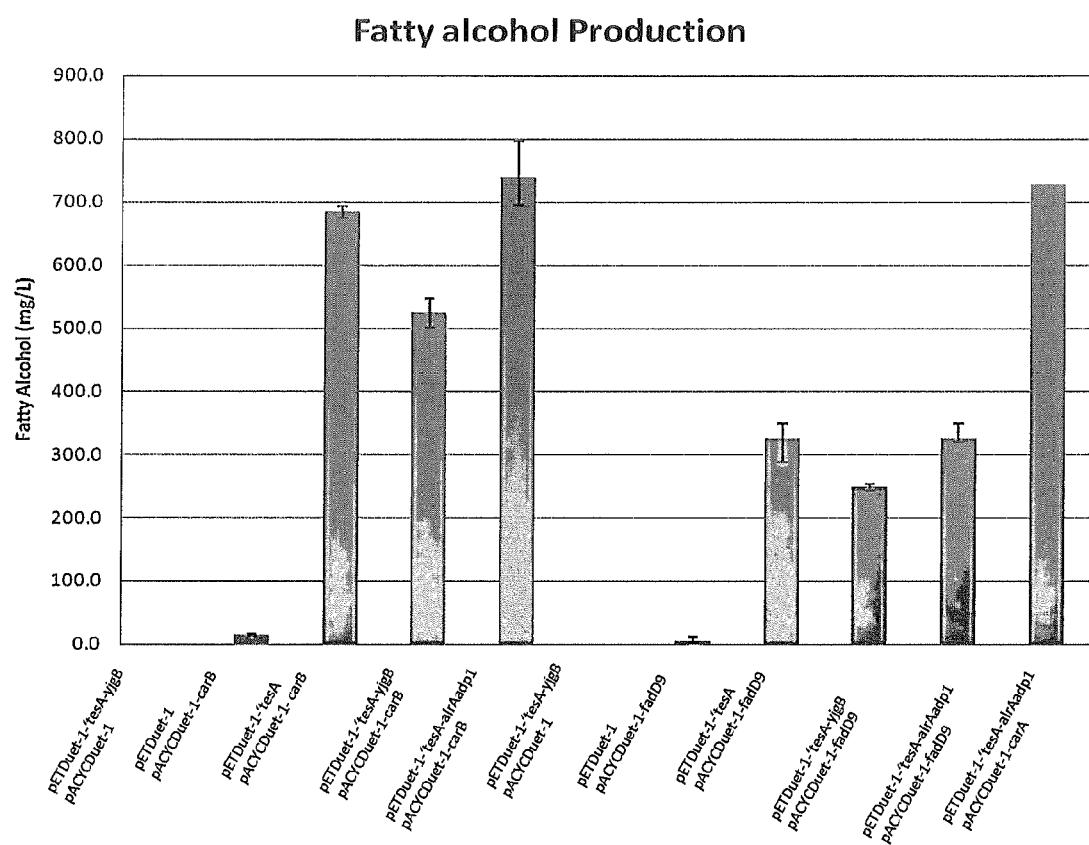
**References Cited**

## OTHER PUBLICATIONS

- Zang et al., "Optimum Conditions for transformation of *Synechocystis* sp. PCC 6803," *The Journal of Microbiology*, Jun. 2007, vol. 45, No. 3, pp. 241-245.
- Zhang et al., "Inhibiting Bacterial Fatty Acid Synthesis", *J.Biol. Chem.* 281(26): 17541-17544 (2006).
- Zhang et al., "Molecular Effect of FadD on the Regulation and Metabolism of Fatty Acid in *Escherichia coli*," *FEMS Microbiol Lett*, 259, 2006, pp. 249-253.
- Zhang et al., "Structural Basis for Catalytic and Inhibitory Mechanisms of β-Hydroxyacyl-acyl Carrier Protein Dehydratase (FabZ)", *J.Biol.Chem.* 283(9):5370-5379 (2008).
- Zhang et al., "The FabR (YijC) Transcription Factor Regulates Unsaturated Fatty Acid Biosynthesis in *Escherichia coli*\*," *The Journal of Biological Chemistry*, vol. 277, No. 18, Issue of May 2002, pp. 15558-15565, 8 pages.
- Zheng et al., "Thioesterase II of *Escherichia coli* Plays an Important Role in 3-Hydroxydecanoic Acid Production," *Applied and Environmental Microbiology*, vol. 70, No. 7, Jul. 2004, pp. 3807-3813, 7 pages.
- Zhu et al., "Functions of the *Clostridium acetobutylicum* FabF and FabZ proteins in unsaturated fatty acid biosynthesis", *BMC Microbiology* 9:119 (2009).
- Zimhony et al., "Characterization of *Mycobacterium smegmatis* Expressing the *Mycobacterium tuberculosis* Fatty Acid Synthase I (fasI) Gene", *J.Bacteriol.* 186(13): 4051-4055 (2004).

- International Search Report and Written Opinion from PCT/US2007/011923, mailed Feb. 22, 2008.
- Partial International Search Report from the ISA of the EPO from PCT/US2008/058788, mailed May 11, 2008, 4 pages.
- International Search Report and Written Opinion of the ISA of the EPO for PCT/US2009/044403, mailed Sep. 25, 2009, 10 pages.
- International Search Report and Written Opinion from PCT/US2009/044409, mailed Jan. 29, 2010.
- International Search Report and Written Opinion from PCT/US2009/59904, mailed Apr. 5, 2010.
- International Search Report and Written Opinion from PCT/US2009/59903, mailed Jun. 2, 2010.
- International Search Report and Written Opinion from PCT/US2010/050026, mailed Jan. 6, 2011.
- Final Office Action on U.S. Appl. No. 12/575,430, mailed Jun. 7, 2012.
- Final Rejection on U.S. Appl. No. 12/575,430, mailed Nov. 29, 2010.
- Final Rejection on U.S. Appl. No. 12/575,430, mailed Sep. 1, 2011.
- Non-Final Office Action on U.S. Appl. No. 12/575,430, mailed Dec. 27, 2011.
- Non-Final Office Action on U.S. Appl. No. 12/575,430, mailed May 13, 2011.
- Non-Final Office Action on U.S. Appl. No. 12/575,430, mailed Jun. 10, 2010.
- Non-Final Office Action on U.S. Appl. No. 12/575,430, mailed Jul. 7, 2014.
- Notice of Allowance on U.S. Appl. No. 12/575,430, mailed Dec. 8, 2014.

\* cited by examiner

**FIG. 1**

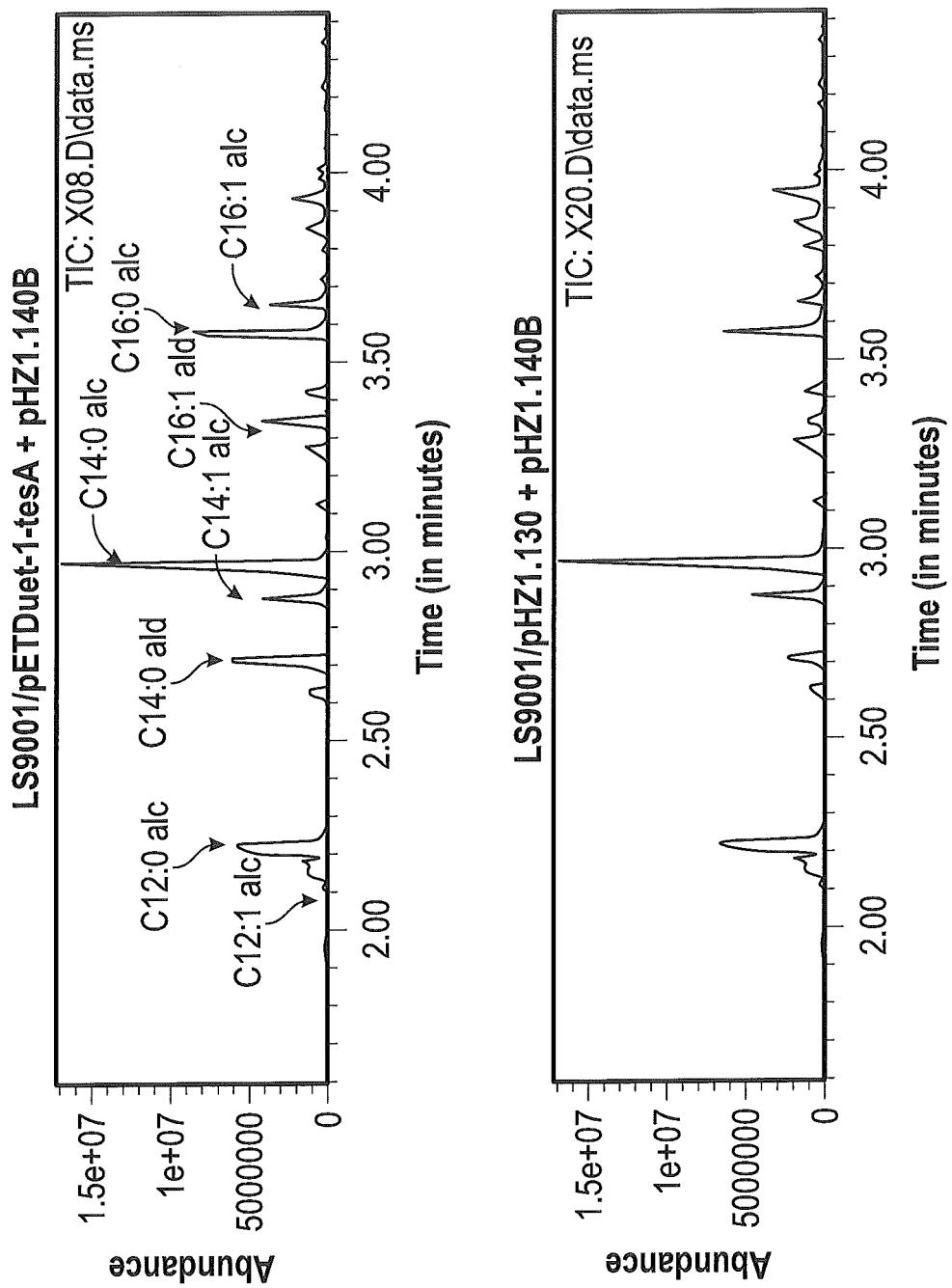


FIG. 2

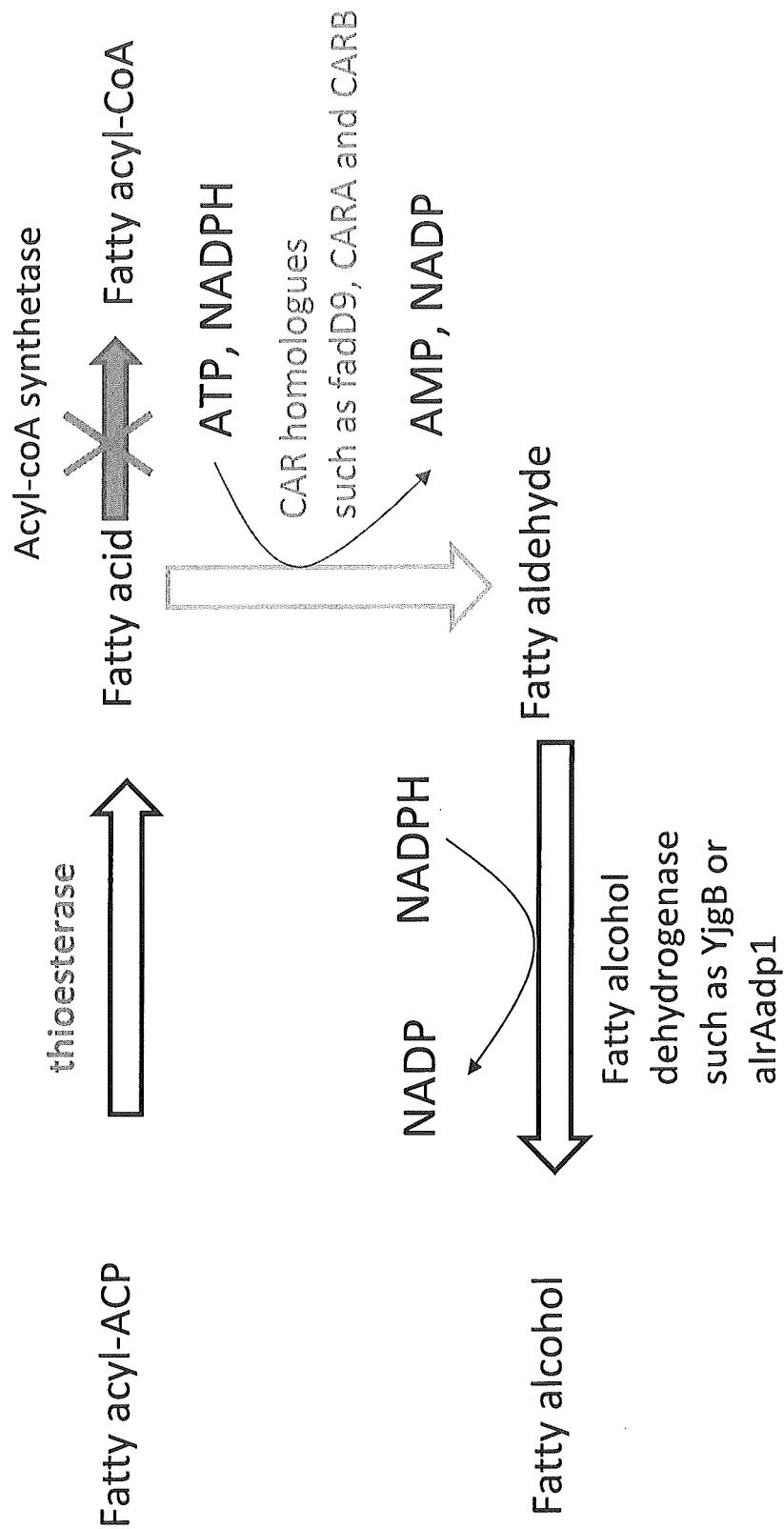


FIG. 3

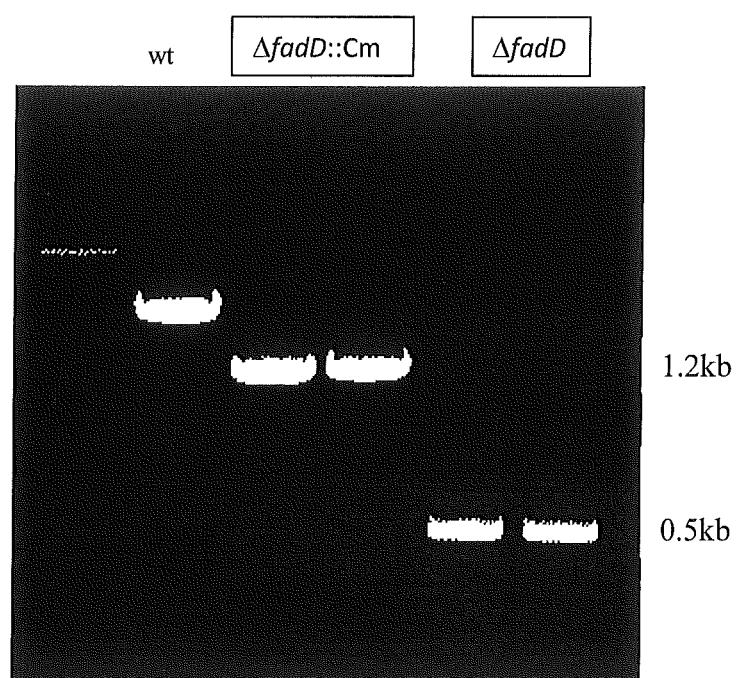


FIG. 4

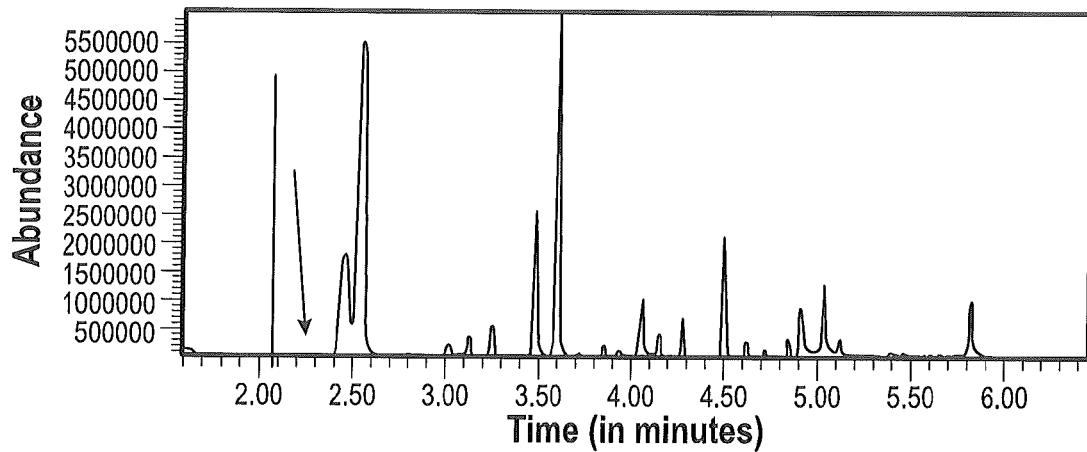


FIG. 5A

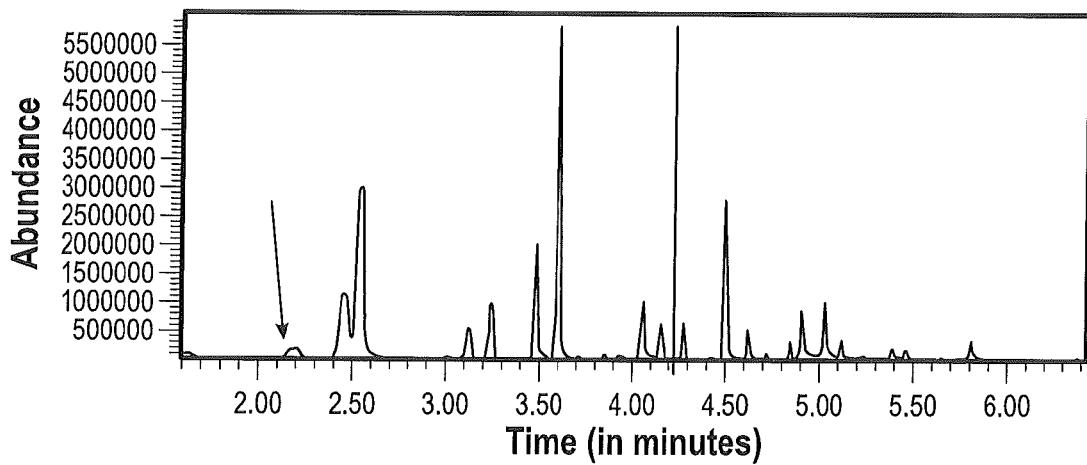


FIG. 5B

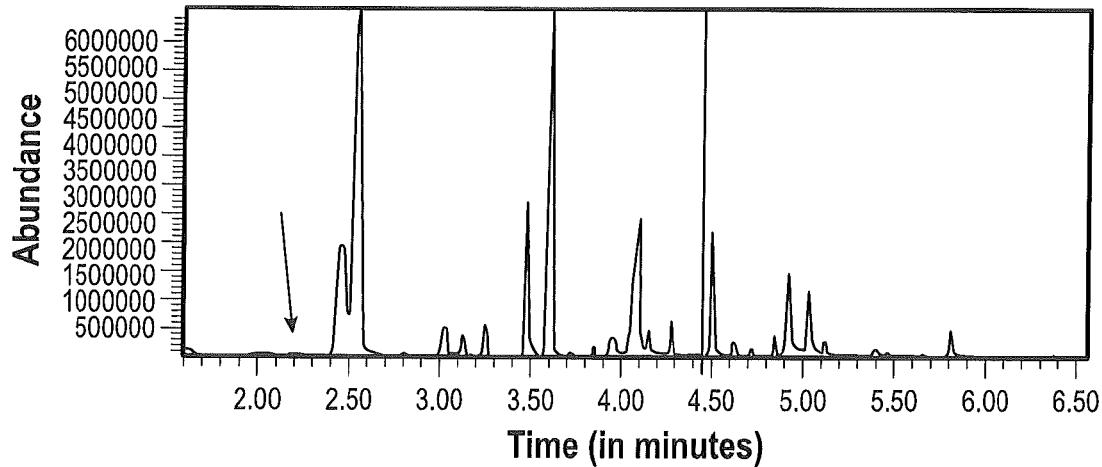


FIG. 5C

FIG. 6A

AAR91681.1

### Nucleotide sequence (SEQ ID NO:15)

>gi|40796034:488-4012 Nocardioides sp. NRRL 5646 ATP/NADPH-dependent carboxylic acid reductase (car) gene, complete cds

ATGGCAGTGGATTACCGGATGAGCGGCTACAGCGCCGCATTGCACAGTTGTCAGAAGATGAGCAGG  
TCAAGGGCGCACGTCCGCTGAAGCGGTGAGCGCGGGTACGCCGACCGCCGGCGCCGGTACGCCGAGAT  
CGCCGCCACTGTTATGGGGTTACGCCGACCGCCGGCGCCGGTACGCCGAGATGCGCTGGCGCAGAT  
GACGACGCACGGCCGACCTCGCTGGTTACTTCCCCGATTCGAGACATCACCTATCGCGAACTG  
GGCAGCGAGTCGGCAGGTTGCCGCGCTGGCATCATGATCCCAGAGAACCCCTGCGCAGGTGATT  
CGTCGCCCTGCTCGCTTACCCAGCATCGACTACGCCACCCCTGACCTGGCGATATCCACCTCGCGCG  
GTTACCGTGCCTGAGGCCAGCGCGGGTGTCCAGCTGATCGTATCTCACCGAGACTTCGCCG  
GGCTGCTGCCCTGACCCCGGAGCACCTCGATGCCGGTCAAGTGCTACTCGCGGCCACCACACCGA  
ACGACTGGTGGTCTCGACTACCACCCCGAGGACGACGACCAGCGTGCAGGCTTCGAATCCGCCGCC  
CGCCTGCCGACGGGGCAGCTTGGTATCGTCGAAACGCTGATGCCGTGCGTGCCTGGGGCGCGACT  
TACCGCCGCGCCACTGTTCCGACACCGACGACGACCCGCTGGCCCTGCTGATCTACACCTCCGG  
CAGCACCGGAACCGCAAGGGCGCATGACACCAATCGGTTGCCGACGATGTCAGGGAACTCG  
ATGCTGAGGGAACTCGAACGGTCTGGATCAATCTCAACTACATGCCATGAGCCACATCGCGGTC  
GCATATCGTGTCCGGCGTCGCTCGCGTGGCACCGCATACTCGCGCCAAGAGCGACATGTCGAC  
ACTGTTGAAGACATCGGTTGGTACGTTCCACCGAGATCTTCTCGTCCCGCGTGTGCGACATGGT  
TTCCAGCGTATCAGAGCGAGCTGGACCGCGCTCGTGGCGGCCACCTGGACACGCTCGATCGGG  
AAGTGAAGCCGACCTCCGGCAGAACTACCTCGGTGGCGCTTCTGGTGGCGGTGTCGGCAGGGCG  
GCTGGCCGCGGAGATGAAGACGTTCATGGAGTCCGTCCTCGATCGCACGACGGTACGGGTCG  
ACCGAGGGCGGGCGCAAGCGTGTGCTGACAACCAGATCCAGCGGCCGCCGTGCTGATTACAAGCTG  
TCGACGTGCCGAACTGGTTACTTCCGACCGACCGGCCGATCCGCGGGTGAAGCTGTTGAAAGGC  
GGAGACACGATTCCGGCTACTACAAGCGCCCGAGGTCACCGCGGAGATCTCGACGAGGACGGCTTC  
TACAAGACCGCGATATCGTGGCCGAGCTCGAGCACGATCGGTGGTCTATGTCGACCGTCGAAACAATG  
TGCTCAAACGTGCGAGGGCGAGTTCGTGACCGTCCGCCCCATCTCGAGGCCGTGTCGCCAGCGCCGCT  
GATCCGGCAGATCTCATCTACGGCAGCGAACGTTCCATCTGTCGCGGTGATCGTCCCCACCGAC  
GACCGCCTGCGCGGCCGACACCGCCACCTTGAATCGGACTGGCGAATCGATTAGCGCATCGCC  
AGGACGCGAACCTCGCAGCCCTACGAGATTCCGCGGATTCCTGATCGAGACCGAGCCGTTACCG  
CAACGGACTGCTCCGGCATCGAAGCTGCTGCCCAACTGAGAACGCTACGGCCTCAGCTG  
GAGCAGATGTACACCGATCTCGCAGAGGCCAGGCCGATGAGCTGCTGCCCTGCCGCGAAGCCGCG  
ACCTGCCGTGCTCGAAACCGTCAGCCGGCAGCGAACAGCGATGCTCGGCGTCCGCGATATGCG  
TCCCGACGCGACTTCACCGACCTGGCGGGCGATTCCCTTCCGCGTGTGTTCTCGAACCTGTCGAC  
GAGATCTCGGGTGTGAGGTGCCGGTGGGTGCTCGTCAGCCCCGGCGAACGAGCTGCCGATCTGGCGA  
ATTACATTGAGGCCGAAACGCAACTCGGGCGCAAGCGTCCCACCTCACCCTCGGTGACGGCGGGTTC  
CGAGATCCGCGCCGCCGATCTGACCCCTCGACAAGTTCATCGATGCCGACCCCTGCCGCCGACAGC  
ATTCCGACGCGCCGGTGCCAGCGCAGACGGTGTGCTGACCGGGCGAACGGCTACCTGCCGTTCC  
TGTGCCCTGAATGGCTGGAGCGGCTGGACAGCGGTGGCACGCTGATCTGCGTGTGCGCGTAGTGA  
CGCGGCCGCGGCCGTAACCGCTGGACTCGCGTTCGACAGCGCGATCCCGCCTGCTCGAGCACTAC  
CAGCAACTGGCGCACGGACCTGGAACTCCCGGCTGATATCGCGACCCGAATCTCGGTCTGGAC  
ACCGCAGTTGGCAGCGGTGGCGAACCGTCGACCTGATGTTCCATCCGCCGCGTGGTCAACCG  
CTTCCCTACACCCAGCTTCCGCCCAATGTCGTCGGCACGCCGAAATCGTCCGGTGGCGATCAGC  
GCGCGGCCGCAAGCGGTACCTACCTGTCGACCGTGGAGTGGCGACCGAGGTCGACCCGGCGAGATTC  
AGGAGGACAGCGACGTCGCGAGATGAGCGCGGTGCCGTCGCGAGAGATTCGCGAACGGCTACGG  
CAACAGCAAGTGGGGGGGGAGGTCTGTCGCGCAAGCACCGATCTGTTGGCTGCCGGTCCGGT  
TTCGCTGGACATGATCTGGCGACAGCGGTACCGGGTCAGCTCAACGTCAGGACGTGTTACCC  
GGCTGATCTCAGCCGGTCCGACCGGCGATCGCGCGTACTCGTTCTACCGAACCGACGCGGACGGCAA

**FIG. 6B**

CCGGCAGCGGGCCCCACTATGACGGCTGCCGGCGGACTTCACGGCGGGCGATCACCGCGCTCGGCATCAAGCCACCGAAGGCTTCCGGACCTACGACGTGCTCAATCCGTACGACCATGGCATCTCCCTCGATGAATTCTCGACTGGCTCGTCGAATCCGGCCACCCGATCCAGCGCATCACCGACTACAGCGACTGGTCCACCGTTTCGAGACGGCGATCCGCGCCTGCCGAAAAGCAACGCCAGGCCTCGGTGCTGCCGTTGCTGGACGCC TACCGCAACCCCTGCCGGCGTCCCGCGCGATACTCCCGCCAAGGAGTTCCAAGCGGCAGTGCAGCAAACAGCCAAATCGGTCCGGAACAGGACATCCCGATTGTCCGCGCCACTGATCGATAAGTACGTCAGCGATCTGGAACTGCTCAGCTGCTCTGA

**Amino acid sequence (SEQ ID NO:16)**

>gi|40796035|gb|AAR91681.1| ATP/NADPH-dependent carboxylic acid reductase [Nocardia sp. NRRL 5646]

MAVDSPDERLQRRIQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAATVMAGYADRPAAGQRAFELNTDDATGRDSLRLLPRFETITYRELWQRVGEVAAWHDPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPIQASAAVSQLIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVF DYHPEDDDQRAAFESARRRLADAGSLVIVETLDRAVRARGRDLPAAPLFVPDTDDDPLALLIYTSGSTGTPKGAMYTNRLAATMWQGNSMLQGNSQRVGGINLNYPMSHIAGRISLFGVLARGGTAYFAAKSDMSTIFEDIGLVRPTEIFFVPRVCDMVFQRYQSELDRRSVAGADLTDREVKADLRQNYLGGRFLVAVVGSAPLAEMKTFMESVLDLPLHDGYGSTEAGASVLLDNQIQRPPVLDYKLVDVPELGYFRTRDRPHPRGELLKAETTI PGYYKRPEVTAEIFDEDGFYKTGDIVAELEHDLRVYVDRRNNVLKLSQGEFVTVALEAVFASSPLIRQIFIYGSERSYLLAVIVPTD DALRGRTDATLKSALAESIQRIAKDANLQPYEIPRDFLIETEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDLATGQADELLALREAADLPVLETVSRAAKAMLGVASADM RPD AHFTDLGGDSLSALSFSNLLHEIFGVEVPVGVVVSPANELRDLANYIEAERNSGAKRPTFTSVHGGSEIRAADLTLDKFIDARTLAAADSIPHAPVPAQTVLLTGANGYLGRFLCLEWLERLDKTGGTLICVV RGSDAAAARKRLDSA FDSDGDPGLLEHYQLAARTLEVLAGDIGDPNLGLDDATWQRLAETVDLIVHPAALVNHVLPYTQLFGPNVVGTAEIVRLAITARRKPVTYLSTVGVADQVDPAEYQEDSDVREMSAVRVVRESYANGYGN SKWAGEVLLREAHDLCLGPVAVFRSDMILAH SRYAGQLNVQDVFTRLILSLVATGIAPYSFYRTDADGNRQRAHYDGLPADFTAAAITALGIQATEGFRTYDVLPYDDGISLDEFVDWLVESGHPIQRITDYS DWFHRFETAIRALPEKQRQASVPLLLDAYRNPCPAVRGAILPAKEFQAAVQTAKIGPEQDIPHLSAPLIDKYVSDLLELLQLI

**FIG. 7A**Motif 1

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); and

-G-X-X-G-X-L-G (SEQ ID NO:8); and

-L/I/G-G-D-S-X-X-A (SEQ ID NO:9); and

-[LIVMFY]-{E}-[VES]-[STG]-[STAG]-G-[ST]-[STEIA]-[SG]-X-[PASLIVM]-[KR] (SEQ ID NO:10), where {X} stands for any amino acid except X and [X<sub>1</sub>X<sub>2</sub>] stands for X<sub>1</sub> or X<sub>2</sub>

Motif 2

RTVLLX<sub>1</sub>GAX<sub>2</sub>GX<sub>3</sub>LGRX<sub>4</sub>LX<sub>5</sub>LX<sub>6</sub>WL (SEQ ID NO:11)

where X<sub>1</sub> is S or T;

X<sub>2</sub> is T or N;

X<sub>3</sub> is F or W;

X<sub>4</sub> is F or Y;

X<sub>5</sub> is A or T; and

X<sub>6</sub> is E or Q

Motif 3

LXXGXXGXLGXXLXLXWLXR (SEQ ID NO:12)

Motif 4

WAXEVLLR (SEQ ID NO:13), where X can be any amino acid; or

LXXGXXGXLGXXLXX<sub>1</sub>XX<sub>2</sub>LX<sub>3</sub>R (SEQ ID NO:14), where

X<sub>1</sub> is Leu or Ile;

X<sub>2</sub> is Trp or Leu; and

FIG. 7B

X<sub>3</sub> varies between 13 amino acids or 14 amino acids

Motif 5

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); and

-L/V/I-G-G-D-S-X-X-A (SEQ ID NO:9); and

-[LIVMFY]-{E}-[VES]-[STG]-[STAG]-G-[ST]-[STEIA]-[SG]-X-[PASLIVM]-[KR] (SEQ ID NO:10), where {X} stands for any amino acid except X and [X<sub>1</sub>X<sub>2</sub>] stands for X<sub>1</sub> or X<sub>2</sub>; and

RTVLLX<sub>1</sub>GAX<sub>2</sub>GX<sub>3</sub>LGRX<sub>4</sub>LX<sub>5</sub>LX<sub>6</sub>WL (SEQ ID NO:11), where

X<sub>1</sub> is S or T;

X<sub>2</sub> is T or N;

X<sub>3</sub> is F or W;

X<sub>4</sub> is F or Y;

X<sub>5</sub> is A or T; and

X<sub>6</sub> is E or Q

**FIG. 8A****NP 217106 (FADD9)****Nucleotide sequence (SEQ ID NO:17)**

>gi|57116681:2917871-2921377 Mycobacterium tuberculosis H37Rv, complete genome

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATAACGCCAGCGACGCCAGTTCGCCG  
CCGCCAGTCCCAACGAGGCATCACCCAGGCATCGACCAGCCGGGGTCGCGCTTCCACAGCTCATCCG  
TATGGTCATGGAGGGCTACGCCGATCGGCCGACTCGGCCAGCGTGCCTCCGCTTCGTACCGAACCCC  
GACAGCGGCCGACCATGGTCGAGCTACTGCCGCGTTCGAGACCATCACCTACCGCGAAGTGGGGCCC  
GCGCCGGCACATTGGCACCGCGTTGAGCGTGAGGCCGCGATCCGGCGGGCGACCGGGTTGCGTGCT  
GGGCTTCAACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGCGCCGTGCGTTCCA  
CTGCAGACCAGTGCGCCGGTACCGGGTTGCGCCCGATCGTACCGAGACCGAGCCGACGATGATGCCA  
CCAGCATCGACAATCTGGCGACGCCGCTCGAAGTGTGCTGGCCGGTACGCCCGGCCGCTGGTCGATT  
CGATTACCACGGAAGGTTGACACCCACCGCGAGGCCGCTCGAAGCCGCCGAGCTCGTTGGCCGGCTCG  
GTGACCATCGACACACTTGGCGAAGTGTGATCGAACCGCGAGGGCGTCCGGCCACACCCATTGCCGACA  
GCGCCGACGACGCGCTGGCGCTGCTGATTACACCTGGGTAGTACCGCGCACCCAAAGGCCGATGTA  
TCGCGAGAGCCAGGTGATGAGCTCTGGCGCAAGTCGAGTGGCTGGTTGAGCCGAGCGGTTACCCCTCG  
ATCACGCTGAACCTCATGCCGATGAGGCCACGTCGGGGGCCGTCAGGTGCTTACGGGACGCTTCCAACG  
GCGGTACCGCCTACTTCGTCGCAAGAGCGACCTGTCGACGCTGTCGAGGACCTCGCCCTGGTGCGGCC  
CACAGAATTGTGTTCTGCGCGCATCTGGACATGGTGTGCGAGAGTCCACAGCGAGGTGCGACCGC  
CGCTTGGTGGACGGCGCCGATCGAGCGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCCGGAGAACGTG  
TCGGCGGACGGTTGTCATGGCGCTGACCGGTTCCGCGCGATCTCGCTGAGATGACGGCTGGTCAACGAC  
GTCCTGCTGGCCGACGTGCAATTGGTGGAGGGTACGGCTCCACCGAGGCCGGGATGGTCTGAACGAC  
GGCATGGTGCAGGCCGGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGCTACTTCGGCA  
CCGATCAGCCCTACCCCCGGGCGAGCTGCTGGTCAAGACGAAACCATGTTCCCGGCTACTACCGCG  
CCCGGATGTCACCGCCGAGGTGTCGACCCGACGGCTTACCGGACCGGGGACATCATGGCAAAGTA  
GGCCCCGACCAGTTCGCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCATCG  
CCGTGCGAAGCTCGAGGCCGGTGTCCGGCGACAGCCGCTGGTCCGACAGATCTCATCTACGGCAACAG  
TGCCCCGGCCTACCCGCTGGCGTGGTTGTCGGCTCCGGGACGCCGCTTCTCGCCATGGCATCGAGAAT  
CTCAAGCCCGTGAATCGCGAGTCCCTCGAGGAGGTAGCGAGGGCGGCCGCTGCAATCCTACGAGATT  
CACCGCACTTCATCATCGAAACCACGCCGTTCACCTGGAGAACGCCGCTGCTCACCGCATCCGAAAGCT  
GGCACGCCGCGAGTGAAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTACCGAGCTGGCGATAGC  
CAATCCAACGAGCTGCGCGAGCTGCCGAAAGCGTCCCGATGCGCCGGTCTCCGACGCTGTGCCGTG  
CCGCGGCTGCGTGTGCTGGCTTACCGCTGCGGATGTGCGGCCGGACGCCACTTCGCGACCTGGTGG  
TGACTCGCTCTGGCGCTGCGTGTGCTGGCAACCTGCTGACGAGATCTCGCGCTGACGTGCCGTGGG  
GTCATGTCAGCCCCGGCAAGGCACCTGGGGCCCTGGCGACCGACATCGAAGCAGCGCAGCCGTC  
GGCGACCCAGCTCGCCTCGATACACGGTGCCTCCGCGACCGAAGTGCACGCCAGCGACCTCACGCTGGA  
CAAGTTCATCGACGCTGCCACCCCTGGCGCAGCCCCGAACCTGCCGACCGAGCGCCAAAGTGCAC  
GTACTGCTGACCGCGCCACCGGTTTTGGGCGCTACCTGGCGCTGGAATGGCTGACCGCATGGACC  
TGGTCAACGGCAAGCTGATCGCTGGCTGGCGCAGATCCGACGAGGAAGCACAAGCCGGCTGGACCG  
GACGTTGAGACGGCGACCCGTTATGGTGCCTGACCGCAATTGGGCGCCGGCCCTCGAGGTG  
CTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCTGGACCGGGTACCTGGCAGCGGCTAGCCGACACGG  
TGGACCTGATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTGCCAAA  
CGCGCGGGCACGCCGAGTTGCTTGGCTGGCGTGCACGCCAGCGCAAGCCATACATCTACACCTCG  
ACGATCGCCGTGGCGAGCAGATCCGCCGGAGCGTTACCGAGGACGCCGACATCCGGCCATCAGCC  
CGACCCGAGGATCGACGACAGCTACGCCAACGCCGACCGAAGCAAGTGGGCCGGGAGGTGCTG  
GCGCGAAGCTACGAGCAGTGCAGGCCGCTGCCGGTGCACGGTCTCCGCTGCCACATGATCCTGCCGACACC  
AGCTATACCGGTCAGCTCAACCTGCCGACATGTTCACCCGGCTGATGCTGAGCCTGGCCGCTACCGG  
TCGCACCCGGTTGCTTCTATGAGCTGGATGCGCACGCCAATCGCAACCGGCCACTATGACGGCTTG  
GGTCAATTGTCGCGAGAAGCCATTGCAACCTGGGACACATAGCCGGACCGTTGTCACCTACAC  
GTGATGAACCCCTACGACGACGGCATCGGCGTGGACGAGTTGCTGCGACTGGCTCAACTCCCCAACTAGCG

**FIG. 8B**

```
GGTCGGTTGCACGATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTGAGACTTCGCTGCC
TGCCTTGCCGGATGCCAGGCCACGCCCTCGCTGCCCTGCTGCACAACACTACCGAGAGCCTGCAAAG
CCGATATGCCGGTCAATGCCGCCACCGACCAGTTCCGCGCTGCCGTCCAAGAAGCGAAATCGGTCCGG
ACAAAGACATTCCGCACCTCACGGCGCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGCT
GCTGTGA
```

**Amino acid sequence (SEQ ID NO:18)**

>gi|15609727|ref|NP\_217106.1| fatty-acid-CoA ligase [Mycobacterium tuberculosis H37Rv]

```
MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRVMMEGYADRPALGQRALRFVTDP
DSGRTMVELLPRFETITYRELWARAGTLATALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVP
LQTSAPVTGLRPIVTETEPTMIATSIDNLGDAVEVLGHAPARLVFDYHGKVDTREAVEAARARLAGS
VTIDTLEELIERGRALPATPIADSADDALALLIYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPSGYP
SITLNFMPSHVGGRQVLYGTLNSGGTAYFVAKSDLSTLFEDDLALVRPTELCFVPRIWDMVFAEFHSEVDR
RLVDGADRAALEAQVKAELRENVLGGRFVMAUTGSAPISAEMTAWVESLLADVHLVEGYGSTEAGMVLD
GMVRRPAVIDYKLVDVPELGYFGTDQPYPRGELLVKQTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMKV
GPDQFVYLDRRNNVLKSQGEFLIAVSKLEAVFGDSPLVRQIFIFIYGN SARAYPLAVVVPSGDALSRHGIEN
LKPVISESLSQEVARAAGLQS YEIPRDFIIETTPFTLENGLLTGIRKLRPQLKKFYGERLERLYTELADS
QSNEIRELRLQSGPDAPVLP TL CAAA ALLGSTAADVRPDAHFADLGGDSL SALSLANLLHEIFGVDPVVG
VIVSPASDLRALADHIEARTGVRRPSFASI HGRSATEVHASDLDFK FIDAATLAAAPNLPAPSAQVRT
VLLTGATGFLGRYLALEWLDRMDLVNGKLI CLVRARSDEEAQARLDATFDSDGPYILVRHYRELGAGRLEV
LAGDKGEADLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALTGKRKPYIYTS
TIAVGEQIPPEAFTEDADIRAIISPTRRIDDSYANGYANSK WAGEVLLREAHEQCGLPVTVFRCDMILADT
SYTGQLNLPDMFTRIMSLAATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAI TLGTHSPDRFVTYH
VMNPYDDGIGLDEFVDWLNSPTSGSCTIQRIADYGEWLQRFETSLRALPDRQRHASLLPLLHN YREPAK
PICGSIAPTDQFRAAVQEAKIGPDKDIPH LTAAIIAKYISNLRLLLG
```

**ABK75684 (CARA)****Nucleotide sequence (SEQ ID NO:19)**

>gi|118168627|3015785-3019291 Mycobacterium smegmatis str. MC2 155,  
complete genome

```
TTACAGCAATCCGAGCATCTGCAGGTTGCTGATGTA CTTGACGATCACGTGGCCGTGACGTGCGGAATG
TCC TTGTCGGGGCCGATCTCGCGTCTGCACCGCGGACCGGATCGGTGGGTGCCATGGCACCGC
ACACGGGC GG TGAGGGCTGCTGATAGTTGTGCAGCAGCGGAGCGAGCCCTGACGTTGCCGTCCGG
CAGGGCCC CGAGTGC GGTTCTGAACCGGCTCAGCCAGGTGGCGTAGTCTGCGACGCGGTGCACGGGTAG
CCGGCCTCGATCAGCCAGTCCACGTACTCGTCGAGGCCGATGGTGGAGATCGCCTCGCGATGAAC TCCACGGGAG
ACGTCTGAATCCGTCGGTGACCTCGCAGGCCGATGGTGGAGATCGCCTCGCGATGAAC TCCACGGGAG
CCCGTCGTAGTGGGCGCGCTGCCGGTTGCCGTCCGCATCGAGTTCGTAGAACGAAACCGGGCGCGATGCCG
GTCGCCACGAGGCTCAGCATCAGGGGGTGAACATGTCCGGCAGGTTCACTGACCCGAGTAGTCGTGT
CGGCCAGGATCATGTCGCAGCGGAACACCGAGACCGGACACCACACCAGTCGTGCGCTCCCGCAGCAG
GACCTCGCCGGCCCACTTGCTGTTGCCGTAGCCGTGGCGTAGAGTCGTGCGACCCGGCGCTCGCGCTG
ATCTCGCGGATGTCGGCGTCTCGACCGAAGCAGCGCTCGGGGGAGATGCCCTGCCCACACCGATCGTCGAGA
CGTACACGTACGGCTTGATCGTGGTGGTCAGCGCGATCCGGATGAGTTCGGCGGTGCCGAGCGCATTGGG
TCCGAACATCTGGCTGACGGCAGGACGTGATTGACCAGGGCGCCGATCGACGATCAGATCGACGGTG
TCGGCCAGTCGCTGCCACGTGTCGTGGTCGAGACCCAGATCGGCCTCGCCCTGTCACCGCGATCACCT
CGAGGTGATCGGCTGCCAGCGCGCGTAGTGTGCGAGCAGTGTGCGTCCCCGGTGTGAAACGTGGCGTC
```

**FIG. 8C**

CAGACGCGCCGGGCTCGTCGCTGCCGCACCAGGAGATCACCTGCCGTCCACCAGGTCC  
 ATCGCCTCCAGCCATTCCAGCGCAGATAGCGGCCAGGAACCCGGTGGCGCCGGTCAGCAGCACGGTGC  
 GGATCTCGGTGCCGAACCGCGCAGACCCGGCGGCCAGGGTCTTGGCGTCAGTAACCTGCCAG  
 GGCAGATCACCGCGCGCACCTCGGTGGCGCAGCGTACCGACCGTATGTGGCGCTTGGAG  
 CCGCGCAGTCGCCCTCGATGTAGGCCCGACGCCCTGCCAGGTGGCGCCGGCTGACGATGACGCCGA  
 CCAGCACGTCGACATCGAAGATCTCGCAACAGGTCGAGAAGCTCAAGGCCACAACGAATCTCCACC  
 CAGATCGGTGAAGTGCGCATCGGACCGCAGATCCGTGACGGAGGCACCGAGCAGTGCACCGCGCG  
 CTGACGGTCTCGACACGGGCCGTGGCTCCGTTGCGGCCAACTCGCGCAACTCGTTGGCTGCCCT  
 CGGCCAGGTGGTGTAGAGCTGTCAGGCCGTAGTGCCTCAGTTGGCCGGGCCAGCTT  
 CGGGATAACGGTCAGCAGGCCCTTCCAGCGTAAAGGTGTTCTCGACGAGGAAGTCACGCCGGATC  
 TCATACGACTGCAATCCGGCGCTCGCCGCTCTGAGTCGCTGATGCGCAGCTTGAGTTCTG  
 CACCGTCCCACGTGACAGTGCCTCTCGGTGGACCACGACCGCCAGCAGATAGGACCGCGCTGTT  
 GCCGTAGACGTAGATCTGGCGTACCAAGGGGTGTCGCCAACACGCCCTCCAGCTTGGAGACCGTGAC  
 AATTGCCCTCGGACAGTTCAAGCACGTTGTCGGCGGTGAGGTATTCGAGATGGTCGGGCCAGCT  
 CGGCCAGATGTCGCCGGTGGTAGTACCCGCTCTCGTAACATCTGGCGGTGATCTCCGACGCTT  
 GTAGTAGCCGGGAAACATCTGCTCGGACTTGACAGAACAGTTGCGCCGGGTAGGGCCGGTCCGTGGC  
 AAGTAGCCGAGATCGGGCACGTCGACAGCTGAGTCGATGACCGGGCGCTGGATCTGCCGTC  
 TGAACACCGCGCCGGCTCGTGGAGCCGATCCGCTCAGCAGATGCACTGCGAGCAGGTCTCGACCCA  
 GCTCTCATCTCCGCCAGATGGGAGCCGATCCGCTCAGGCCGAAACGAATCGCCGCCGAGCAGTTGG  
 GTGCGGACCTCTCGAGGACTCGGGCTTCGGCTCGTGGATCOCTCGGATCCCTCGCCGGCGTGGAGG  
 GGCTCTGGTACTCTGGAACAGCATGTCAGGAGATGCGAGGAACGAAGTTGAGCTGCGTGGCCGACGAG  
 GGCAGGGTCTCCAGGAAGGTGGACAGGTGCGTGCAGTACCGGTTCCGCCGCTGGCAGT  
 GTGCTGCACAGGATGCCGCCCATGACGTGACTCATGGCATGAAGTTCAAGGTGATGACGGCATCA  
 CGCCGAGGGTCTCGTCCCACCGGGCTTGGACCCGGCTGCCACATCGTGGCGGTCTGGACTCGGGGTA  
 CATCGGCCCTGGGAGTGCCGGTCTGCCGGAGGTGAGATGAGAAGGGTCAGCGGGTCCGCTCGT  
 GGCACGTAGAGCGGTGCGTCGGCGAGTGACCGCCGCTCCAGTGCCTCGTGTACGCTCGACGACA  
 CGCCGGTCCGAGCTTGCCTGCCGCTCGAACGCCACGCTCACGCTGATCGTGCACCTCGTGGCTGTA  
 GTCGAACACCACAGTCGCGACGGCGGGCCGGACTCGACGAGAGCGACTGCGTGGCGAGGAAGTC  
 ACGCTCGACGCGATCACCTGGCTCGGTCTCGCGACGATCGCTGCAAGTTGGGCCACCGGGCAGCTGG  
 TCTGCAGCGGTACGGACACGGCGCCGAGTTGAGCAGGGCAGTCGATCGTGTAGTCGACACTGGT  
 GAAACCCAGGATGCCACGCCATTACCGGATGGTTGTCGAGGATTGGTCAACGCCCTGG  
 ATCCGGCCTCGGAGCTGACGGTAGGTGATGGTGTGCAAGCGGGGAGGAGCTTCGCGGTGGTCCG  
 CTTCGTCGGTAGCGAACTCGACGGCGCTTGCAGCGCAGGGCGTCCGATAGCCGCCAGAATCTG  
 TTTGACCGCGGAGGAAGGCCAACCTCGGATCGCGCAGCCGCGTGTACGCTCGTGGACGGCG  
 GCGCGAACACTCGGGTCGGTTGAACAAGGGTCAATGCCGGTTGAAGCGGTCTCGCGCTTCGA  
 TCGTCAT

**Amino acid sequence (SEQ ID NO:20)**

>gi|118174788|gb|ABK75684.1| NAD dependent epimerase/dehydratase family protein [Mycobacterium smegmatis str. MC2 155]

MTIETREDRFNRRIDHLFEDPQFAAARPDEAI SAAAADPELRLPAAVKQILAGYADRPA  
 LGKRAVEFVT DEEGRRTAKLLPRFDITIYRQLAGRIQAVTNAWHNHPVNAGDRV  
 AILGFTSDYTTIDIALLELGAVSVP LQTSAPVAQLQPIVAETEPKVIASSVDFL  
 ADAVALVESGPAPSRLVVFDYSH  
 EVDQREAFEAAGKLAG TGVV  
 VETITDALDRGRSLADAPLYVPDEADPLT  
 LIYTSGSTGTPKGAMYPESKTATMWQAGSKARDET  
 LGVMP  
 MSITLN  
 FMPMSHVMGRGILC  
 CSTLASGGTAYFAARS  
 DLS  
 FLED  
 LALVRPTQLNF  
 FPRIWDMLFQ  
 Q  
 EY  
 QSRLDN  
 RRAEGSE  
 DRAEA  
 AVLE  
 VRTQ  
 LLG  
 GRF  
 VSALT  
 GS  
 API  
 SAEM  
 KS  
 W  
 VED  
 LLD  
 MHL  
 LEGY  
 GSTEAGA  
 VFID  
 GQIQR  
 RP  
 PVIDYKL  
 DV  
 PDLGY  
 F  
 ATDR  
 PY  
 PR  
 GELL  
 VK  
 SE  
 QM  
 FPG  
 Y  
 KR  
 PE  
 ITA  
 EM  
 F  
 DED  
 GYY  
 RTG  
 D  
 I  
 VAE  
 LG  
 PD  
 H  
 LEY  
 LD  
 RR  
 NN  
 VL  
 KLS  
 SQ  
 GE  
 F  
 VT  
 SK  
 LE  
 AV  
 FG  
 DS  
 PL  
 VR  
 QI  
 Y  
 VY  
 GN  
 S  
 AR  
 SY  
 LL  
 AV  
 V  
 PT  
 E  
 E  
 A  
 L  
 SR  
 W  
 DG  
 DEL  
 K  
 SR  
 IS  
 D  
 S  
 LQ  
 DA  
 A  
 RA  
 AG  
 LQ  
 SY  
 E  
 I  
 PR  
 DF  
 L  
 V  
 ET  
 T  
 P  
 F  
 T  
 L  
 EN  
 GL  
 LT  
 G  
 IR  
 K  
 L  
 A  
 R  
 P  
 K  
 L  
 K  
 A  
 H  
 Y  
 G  
 E  
 R  
 L  
 E  
 Q  
 LY  
 TD  
 LA  
 EG  
 Q  
 A  
 N  
 E  
 L  
 R  
 R  
 N  
 G  
 A  
 D  
 R  
 P  
 V  
 V  
 E  
 T  
 V  
 S  
 R  
 A  
 A  
 V  
 A  
 V  
 L  
 G  
 A  
 S  
 V  
 T  
 D  
 L  
 R  
 S  
 D  
 A  
 H  
 F  
 T  
 D  
 L  
 G  
 G  
 D  
 S  
 L  
 A  
 S  
 F  
 S  
 N  
 L  
 H  
 E  
 I  
 F  
 D  
 V  
 D  
 V  
 P  
 V  
 G  
 V  
 I  
 V  
 S  
 P  
 A  
 T  
 D  
 L  
 A  
 G  
 V  
 A  
 A  
 Y  
 I  
 E  
 G  
 E  
 L  
 R  
 G  
 S  
 K  
 R  
 P  
 T  
 Y  
 A  
 S  
 V  
 H  
 G  
 R  
 D  
 A  
 T  
 E  
 V  
 R  
 A  
 R  
 D  
 L  
 A  
 G  
 K  
 F  
 I  
 D  
 A  
 K  
 T  
 L  
 S  
 A  
 P  
 G  
 L  
 P  
 R  
 S  
 G  
 T

**FIG. 8D**

EIRTVLLTGATGFLGRYLALEWLERMDLVDGKVICLVRARSDEARARLDATFDTGDATLLEHYRALAAD  
HLEVIAGDKGEADLGLDHDTWQRЛАDTVDLIVDPAALVNHVLПYSQMFГPNALGTAEЛIRIAЛTTIKPY  
VYVSTIGVGQGИSPEAFVEDADIREISATRRVDDSYANGYGNSKWAGEVLLREAHWCGLPVSVFRCDMI  
LADTTYSQQLNLPDMFTRLMLSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAEAISTIGSQVTDGF  
ETFHVMNPYDDGIGLDEYVDWLIEAGYPVHRVDDYATWLSRFETALRALPERQRQASLLPLLHNYQQPSP  
PVCGAMAPTDREFRAAVQDAKIGPDKDIPHVTADVIVKYISNLQMLGLL

**YP 889972 (CARB)****Nucleotide sequence (SEQ ID NO:21)**

>gi|118467340:5821317-5824838 Mycobacterium smegmatis str. MC2 155,  
complete genome

ATGACCAGCGATGTTCACGACGCCACAGACGGCCTACCGAAACCGCACTCGACGACGAGCAGTCGACCC  
GCCGCATGCCGAGCTGTACGCCACCGATCCCGAGTTCGCCGCGCCGCCACCGTTGCCGCCGTGGTCGA  
CGCGCGCACAACCCGGCTGCCGCTGGCAGAGATCCTGCAGACCCCTGTTCACCGGCTACGGTGACCGC  
CCGGCGCTGGATACCGCCCGTGAACCTGGCCACCGACGAGGGCGGGCGACCGTGACCGTCTGCTGC  
CGCGTTCGACACCCCTCACCTACGCCAGGTGTGGTCGCGCTGCAAGCGGTGCCGCCGCCCCCTGCC  
CAACTTCGCGCAGCGATCTACCCCGGCACGCCGTGCGACGATCGGTTCGCGAGTCCCGATTACCTG  
ACGCTGGATCTCGTATGCCCTACCTGGGCTCGTAGGTGTTCCGCTGAGCACAACGCACCGGTACGCC  
GGCTGCCCGATCCTGGCGAGGTGCAACCGCGGATCCTCACCGTGAGCGCCGAATACCTGACCTCG  
AGTCGAATCGTGCAGCTCAACTCGGTGTCGAGCTCGTGGTGTGACCATCACCCGAGGTGAC  
GACCACCGGAGCGACTGGCCCGCGCGTGAACAACACTGCCGGCAAGGGCATGCCGTACCCACCTGG  
ACCGATGCCGACGAGGGCGCCGGGCTGCCGGCGAACCGATCTACACCACCGGACCATGATCAGCGC  
CGCGATGATCCTGTACACCTGGGTTCCACCGCGCACCCAAAGGTGCGATGTACACCAGGGGATGGT  
GCGCGCTGTGGACCATGTCGTTCATCACGGGTGACCCCACGCCGGTCAACGTCAACTCATGCC  
TCAACCACCTGGCGGGCGATCCCCATTTCACCGCCGTGAGAACCGTGGAACCGAGTTACTCGTACC  
GGAATCCGACATGTCACCGCTGTCGAGGATCTCGCGCTGGTGCACCGAACCTGGCCTGGTCC  
CGCGTCGCCGACATGCTTACCGACCCACCTGCCACCGTGCACCGCCTGGTACCGAGGGCGCCACG  
AACTGACCGCCGAGAAGCAGGGCGGTGCCGAACTCGCTGAGCAGGTGCTCGCGGAGCGTGTACCGG  
ATTCTGTCAGCACCGCACCGCTGGCCCGGGAGATGAGGGCGTTCTCGACATCACCGTGGCGCACACATC  
GTCGACGGCTACGGGCTACCGAGACCGCGCCGTGACACCGTGTGATCGTGCAGGGCACCGGTGA  
TCGACTACAAGCTGATCGACGTTCCCGACTCGCTACTTCAGCACCGACAAGCCCTACCCGCGTGGCG  
ACTGCTGGTCAGGTCGCAAACGCTGACTCCGGGTACTACAAGCGCCCGAGGTACCGCGAGCGTCTTC  
GACCGGGACGGCTACTACCACACCGCGACGTATGGCGAGACCGCACCCGACCACCTGGTGTACGTGG  
ACCGTCGCAACAAACGTCCTCAAACACTCGCGCAGGGCGAGTCTGTGGCGGTGCCAACCTGGAGGGTGGT  
CTCCGGCGGGCGCTGGTCCCGCAGATCTTCGTTGACGGCACAGCGAGCGCAGTTCCCTCTGGCGT  
GTGGTCCCACGCCGGAGGGCGCTCGAGCAGTACGATCCGGCGCGCTCAAGGCCGCGTGGCGACTCG  
TGCAGCGCACCGCACCGCACCGGAACCTGCAATCCTACGAGGTGCCGGCGATTTCATCGTCAAGACCG  
GCCGTTCAGCGCCGCAACGGGCTGCTGTCGGGTGCGAAAACGCTGCGGCCAACCTCAAAGACCG  
TACGGGCAGCGCTGGAGGAGATGTAACGCCGATATCGCGGCCACCGCAGGCCAACAGTGTGCGGA  
GGCGCGCGGCCACACACCGGTGATCGACACCCCTACCCAGGCCGCTGCCACGATCCTCGCACCG  
GAGCGAGGTGGCATCCGACGCCACTTCACCGACCTGGCGGGGATTCCCTGTCGGCGCTGACACTTC  
AACCTGCTGAGCGATTCTCGTTGCAAGTTCCCGTCCGACCGAACCGGCCACCAACCTCG  
CCCAACTCGCCCAGCACATCGAGGCGAGCGCACCGCGGGTGACCGCAGGCCAGTTCACCCACCGT  
CGCGCGGAGCCACCGAGATCCGGCGAGTGAGCTGACCCCTGGACAAGTTCATCGACGCCAACGCT  
CGGGCGCACCGGTCTGCCAACGGTACCGACCGAGCCACGCCAGGTGTTGCTCTGGCGCCAACGGCT  
GGCTGGGCCGGTCTCTCACGTTGCACTGGCTGGACGCCGACCGCTGCGACCTGTCGGCGCACCTCAT  
CGTGCAGGCCGGCGACGCCGGCGGCCGACCGCTGACCCAGGCCACCGGATCCGAGTT  
TCCCGCCGTTGCCGAGCTGGCGACCGCCACCTGCGGGTGGTCGCGGGTGACATCGGCGACCGAATC  
TGGGCTCACACCGAGATCTGGCACCCGCTGCCGCCGAGGTGACCTGGTGTGATCCGGCAGCGCT  
GGTCAACCACGTCGCTCCCTACCGGCAGCTGTCGGCCCCAACGTCGTCGGCACGGCGAGGTGATCAAG

**FIG. 8E**

```

CTGGCCCTCACCGAACGGATCAAGCCGTACGTACCTGTCACCGTGTGGCATGGGATCCCCG
ACTTCGAGGAGGACGGCGACATCCGGACCGTGAGCCCGGTGCGCCGCTCGACGGCGGATA
CGCCAACGGCTACGGCAACAGCAAGTGGCCGGCGAGGTGCTGCTGCGGGAGGCCACGATCT
GTGCGGGCTGCCGTGCGACGTTCCGCTCGGACATGATCCTGGCGATCCGCGCTACC
CGCGTCAGGTCAACGTGCCAGACATGTTACACCGCAGACGGGATCTCCCTGGATGTG
TCACCGCAGTCCGAGGCTCTTGATCACCGCGTCGCCCGCGTCGTTCTACATCGGAGACGGTGA
GCGCCCAGGGCGCACTACCCGGCTGACGGTCAATTCTGTTGGCCGAGGCGGTACGACGCT
CGCGCAGCAGCGAGGGATACTGTCCTACGACGTATGAACCCGACGACGGGATCTCCCTGG
ATGTGTCGAGACTGGCAGTCCGGCCGGCATCCGATCGACGGGTCGACGACTACGACGACT
GGGTGCGTCGTTCGAGACCGCGTTGACCGCCTCCCGAGAACGCGCGCACAGACCGTACT
GCCGCTGCTGCACGCGTTCCGCTCCCGCAGGCACCGTTGCGCGCGCACCGAACCCACGG
AGGTGTTCCACGCCGCGTGCACGCGACGAGGCGCTGATCGACAAGTACATACGCGATCT
GCGTGAGTCGGTCTGATCTGA

```

**Amino acid sequence (SEQ ID NO:22)**

>gi|118469671|ref|YP\_889972.1| putative long-chain fatty-acid--CoA ligase [Mycobacterium smegmatis str. MC2 155]

```

MTSDVHDATDGVETALDDEQSTRRIAELEYATDPFAAAAPLPAAVKPGLRLAEILQLTFTGYGDR
PALGYRARELATDEGGRTVTRLLPRFDLTYAQWWSRVQAVAAALRHNAQPIYPGDAVATIGFASPDYL
TLDLVCAVLGLSVPLQHNAVPVSRLAPIAEVEPRIITVSAEYLDLAVESVRDVNSSQLVVFDHHPEVD
DHRDALARAREQLAGKGIAVTTLDAIADEGAGLPAEPIYTADHDQRILAMILYTSGSTGAPKGAMYTEAMV
ARLWTMSFITGDPTPVINVNFMPLNHLGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVP
RVADMILYQHHLATVDRLVTQGADELTAEKQAGAELREQVLGGRVITGFVSTAPLAEMRAFLDITLGAHI
VDGYGLTETGAVTRDGVIVRPPVIDYKLIDVPELGYFSTDKPYPRGEELLVRQSQTLPGYKRPEVTASVF
DRDGYHYTGDVMAETAPDHLYVDRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYGNERSFLLAV
VVPTPEALEQYDPAALKAAALADSLQRTARDAAELQSYEVPAFIVETEPFSAANGLLSGVGKLLRPNLKDR
YGQRLEQMYADIAATQANQLRELRRAAATQPVIDLTQAAATILGTGSEVASDAHTDLGGDSLSALTLS
NLLSDFFGFEVPGTIVNPATNLAQLAQHIEAQRAGDRRPSFTTVHGADATEIRASELTLDKFIDAETL
RAAPGLPKVTTEPRTVLLSGANGWLGRFLTQLWLERLAPVGGLTITIVRGRDDAARARLTQAYDTDPEL
SRRFAELADRHLRVVAGDIGDPNLGLTPEIWHRLAAEVLDVHPAALVNHLVPRQLFGPNVVGTAEVIK
LALTERIKPVTYLSTVSVAMGIPDFEEDDIRTVSPVRPLDGGYANGYGN SKWAGEVLLREAHLCGLPV
ATFRSDMILAHPRYRGQNVPMFTRLLSLLITGVAPRSFYIGDGERFRAHYPGLTVDFVAEAVTLGA
QREGYVSYDVMNP HDDGISLDVFDWLIRAGHPIDRVDDWVRRFETALTALPEKRRAQTVPLLHA
FRAPQAPLRGAPEPEVFHAAVRTAKVGPDIPLHDEALIDKYIRDREFGLI

```

**YP 905678.1****Nucleotide sequence (SEQ ID NO:23)**

>uniprot|A0PPD8|A0PPD8\_MYCUA Fatty-acid-CoA ligase FadD9

```

ATGTCGCCAATCACCGGTAAAGAGCGGGCTCGAGCGCCGCATCCAGGACCTCTACGCCAAC
GACCGCAGTCGCCGCCAAACCGTCACGGCATACCGCAGCAATCGAGCGCCG
GGTCTACCGCTACCCAGATCATCGAGACCGTCATGACCGGATAACGCCGATCGGCCG
CTCGCTCAGCGCTCGGTGAATTCTGACCGATGCCGGCACACCGCTGCCGA
CTGCTCCCCACTTCGAAACCATCAGCTACGGCAGCTTGGGACCGCATCGCGACTG
GCCGACGTGCTCAGCACCGAACAGACGGTAAACCGAGCGACCGGGCTGCTTGGG
TTCAACAGCGTCACTACGCCACGATCGACATGACTTGGCCGGCTGGCGCGGGCT
GTACCACTGCAGACCGAGCGCGGCGATAACCCAGCTGCAGCGGATCGTCGCCGAGACCCAG
CCCACCATGATCGCGGCCAGCGTCGACGCACTCGCTGACGCCACCGAATTGGCTCTGTCC

```

FIG. 8F

GGTCAGACCGCTACCCGAGTCCTGGTGTTCGACCACCACCGCAGGTTGACGCACACCGC  
GCAGCGGTCGAATCCGCCCCGGAGCGCCTGGCTGGCTCGCGGTGCGAAACCCCTGGCC  
GAGGCCATCGCGCGGGCGACGTGCCCCCGGGTGCCTCGCCGGCTCGGCCGCCGGCACC  
GATGTGTCCGACGACTCGCTCGCTACTGATCTACACCTCGGGCAGCACCGGTGCGCCC  
AAGGGCGCGATGTACCCCCGACGCAACGTTGCGACCTCTGGCGAAGGCCACCTGGTTC  
GAAGGCCTACGAGCCGTCGATCACGCTGAACCTCATGCCAATGAGCCACGTATGGC  
CGCAAATCTGTACGGCACGCTGTGCAATGGCGCACCGCCTACTTCGTGGTGAAAAGC  
GATCTCTCACCTTGTTCGAAGACCTGGCGCTGGTGCAGGCCCACCGAGCTGACCTTCGTG  
CCCGCGTGTGGGACATGGTGTGACGAGTTCAGAGTGAGGTGACCGCCGCTGGTC  
GACGGCGCCGACCGGGTCGCGCTCGAAGCCCAGGTCAAGGCCGAGATA CGCAACGACGTG  
CTCGGTGGACGGTATA CAGCGCACTGACCGGCTCCGCCCCGATCTCCGACGAGATGAAG  
GCGTGGGTGAGGAGCTGCTGACATGCATCTGGTGCAGGGCTACGGCTCCACCGAGGCC  
GGGATGATCCTGTACGGAGCCATTGCGCCGGCTACTCGACTACAAGCTGGTC  
GATGTTCCGACCTGGTTACTTCCTGACCGACCGGCCACATCCGCGGGCGAGTTGCTG  
GTCAAGACCGATAGTTGTTCCGGGCTACTACCGAGCGAACGTCAACCGCAGGTG  
TTCGATGCTGACGGCTTCTACCGGACCGGGCACATCATGGCGAGGTGCGGGGGAAACAG  
TTCGTGTACCTCGACCGCCACAACAGTGTGAGCTGTCGAGGGCGAGTTGCTCACC  
GTCTCAAACACTCGAACAGCGGTGTTGGCGACAGCCCACGGTACGGCAGATCTACATCTAC  
GGCAACAGCGCCCGTGCCTACCTGTTGGCGGTGATCGTCCCCACCCAGGAGGCGCTGGAC  
GCCGTGCCTGTCGAGGAGCTCAAGCGCGCTGGCGACTCGCTGCAAGAGGTGCAAAG  
GCCGCCGGCCTGCAAGTACGAGATCCCGCGGACTTCATCATCGAAACAAACACCATGG  
ACGCTGCAGAACAGGCTGCTACCGGCATCGCAAGTTGGCCAGGCCAGCTGAAAAAG  
CATTACGGCGAGCTCTCGAGCAGATCTACACGGACCTGGCACACGGCCAGGCCAGCAA  
CTGCGCTCGCTCGGCCAAAGCGGTGCGATGCGCCGGTGTGGTGCAGGTGCGCG  
GCCGCCGGCGCTGTTGGCGGCAGCGCCTCTGACGTCCAGCCCAGTGCACCTCACC  
TTGGGCCGGCAGTCGCTGTCGGCGCTGTCGTTACCAACCTGCTGCACGAGATCTCGAC  
ATCGATGTCGGCTGGCGTACGAGCTCGCAAAACCGCTCGTACGACCGACCTTCG  
TACGTGAGGCGGCTCGCAAACCGCTCGTACGACCGACCTTCGCGCTCGTCCACGG  
GCCTCGAAATGAGCAGGTCAACCGAGGTGCATGCCGTGACCTGTCCTGGACAAATT  
CATC GATGCCGAAACCCCTGGCGAAGCTCCCGCTGCCGGCAAAACACCCAGTGC  
GTGCTGCTGACCGGGGCCACCGGCTTCTCGGGCGTACCTGGCCCTGGAATGGCTGGAG  
CGGATGGACCTGGTCGACGGCAAACGACTGATCTGCGCTGGCCAGTCCGACACCGAA  
GCACGGCGCGCTGAAAGACGTTGACAGCGGCCCGAAGTCTGGCCACTAC  
CGCGCACTGGCGCGACCGACCTCGAGGTGCTCGCCGGTGACAAGGGCGAGCC  
GGACTGGACCGGCAGACCTGGCAACGCGCTGGCGACACGGTCGACCTGATCG  
GCCGCCCTGGTCAACCACGTACTGCCATACAGCCAGCTGTTGGGCCAACCG  
ACCGCGAGCTGCTGCCGCTCGCGTACCTCCAAGATCAAGCCCTACAGCTACAC  
ACAATCGGTGTCGCCGACCGAGATCCCGCCGCTGGCGTACCGAGGACGCC  
GTCATCAGCGCCACCGCGCGGTGACGACAGCTACGCCAATGGCTATT  
CGCAACAGCAAG TGGCCGGCGAGGTGCTGTCGCGAGGGCGCATGTC  
TTCGCTGCGACATGATCTGGCGACACCACATGGCGGGGACAGCTCAAC  
GTGTTCACCGTATGATCTGAGCCTGGCGCCACGGGTATCGCG  
GAGCTTGGCGCGACGGCGCCGGCAACCGCCACTATGACGGTCTGCC  
ATCGCCGAGGCATTTCGACTTGGGTGCCAGAGCCAGGATGGT  
TCCGGTTGCCCATCAGCGCATCGCTGACTATGGCGACTGGCTGC  
GCACTGCGCGACTGCCGATCGCAGCGCACAGCT  
TATCGGCAGCCGGAGCGGCCGTCGCGGGTCGATGCC  
GCAGGTGCAAGAGGCCAAGATGCC  
ATCGTGAAGTACGTCAGCGACCTGCGCCTACTCGGC  
CTGCTCTGA

**FIG. 8G****Amino acid sequence (SEQ ID NO:24)**

>uniprot|AOPPD8|AOPPD8\_MYCUA Fatty-acid-CoA ligase FadD9

MSPITREERLERRIQDLYANDPQFAAKPVTAITAALAIERPGLPLPQIIET  
VMTGYADRPALAQRSVEFVTDAAGTGHTLRLLPHFETISYGELWDRISAL  
ADVLSTEQTVKPSPDRVCLLGFNSVDYATIDMTLARLGAVAVPLQTSAAIT  
QLQPIVAETQPTMIAASVDALADATELALSGQTATRVLVFDDHHRQVDAHR  
AAVESARERLAGSAVVEETLAEAIARGDVPRGASAGSAPGTDVSDDSLLALL  
IYTGSTGAPKGAMYPRRNVATFWRKRTWFEGGYEPSITLNFMPPMSHVMG  
RQILYGTLCNGGTAYFVVKSDSLSTLFEDLALVRPTELTFVPRVWDMVDE  
FQSEVDRRLVDGADRALEAQVKAEIFRNDVLGGRTSALTGSAPISDEMKG  
AWVEELLDMHLVEGYGSTEAGMILIDGAIRPAVLDYKLVDVPDLGYFLT  
DRPHPRGELLVKTDLSLFPGYQQRAEVTDADFDADGFYRTGDIMAEVGPEQ  
FVYLDRRNNVLKLSQGEFVTVKLEAVFGDSPLVRQIYIYGN SARAYLLA  
VIVPTQEALDAVPEELKARLGDSLQEVAKAAGLQS YEIPRDFIIETTPW  
TLQNLGTTGIRKLRPQLKKHYGELLEQIYTDLAHQADELRSLSRQSGAD  
APVLVTVCRAAAALLGGSASDVPDAHFTDLGGDSLSALSFTNLLHEIFD  
IDVPVGIVVSPANDLQALADYVEAARKPGSSRPTFASVHCASNEQVTEVH  
AGDLSLDKFIDAATLAEAPRLPAANTQVRTVLLTGATGFLGRYLALEWLE  
RMDLVDGKLIKLCRAKSDTEARARLEKTFDGAPPELLAHYRALAGDHLEV  
LAGDKGEADLGLDROTQRLADTVLDLIVDPAALVNHVLPYSQLFGPNALG  
TAELLRLALTSKIKPYSYTSTIGVADQIPPSAFTEDADIRVISATRAVDD  
SYANGYSNSKWA GEVLLREAHVLCGLPVA FRCDMILADTTWAGQLNVPD  
MFTRMILSLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAISTLGA  
QSQDGFBTHYHVMNPYDDGIGLDEFVDWLNESGCPIQRIADYGDWLQRFET  
ALRALPDRQRHSLLPLLHNYRQPERPVRGSIAPTDRFRAAVQEAKIGPD  
KDIPHVGAPIIVKYVSDLRLLLGLL

**ZP 04027864****Nucleotide sequence (SEQ ID NO:25)**

>gi|227980074:564806-568123 Tsukamurella paurometabola DSM 20162  
TpauDRAFT\_4083016\_Cont3, whole genome shotgun sequence

ATGTCGATTGAGACGGTGCAGAACGGCGTCCCCCGCAGAGGGCTCGGTGCCCGCCGACCAGCACCGA  
GCGACTGCCGAGGTGATGCCAGGATCTCGCCAGTCGCGGATCGTCCGGCTCGCGACCCGCGAGG  
CGGGGCCGGGACCCCTACGCCACCGCTCCTATCGGGAGATCTGGCGCGCTCACCGCCTGGTGGCC  
TCCTGGCAGAGCGAAGTGGCTCCGGAGACTTCGTCGCCATCCTCGGCTCACCA GCTCGGACTTCGTGAC  
CGTCGACCTCGCACCACACTGCTCGGCCCGAACGTGCCGCTCAGGCCGGGCCCCCGCGCTCGCA  
TCGC GACC ATCCTCGATGAGACCCGGCGAAGATCCTCGCCGTGAGTGCCGATCAGGTGACCTCGCCCAG  
GAGGCCTGGCCAGTCCGCGGCTACCCCGCGGTGGTGTCTCGACGGCGAACCGCAGGGTACGAGGG  
CATCGAGGGCGACATCCTTCCGGCTCCGCCCTGCCGGCACCGGAGTTCTTCGCGCCCGAGCCGGCACCG  
ATCCTCTCGTCACGCTCATCTACACATCCGGCAGCACCGGTACCCCGAAGGGGGCATGTACACCGAGCAG  
TTGGTTCGCGATGCCGGCTCAAGGTGGACAGCATCGTCGACATCGACATGCCGGCGAGTCGCTCTGCA  
CTTCCGCTATGAGCCATATGTACGGCGCAACTGGCTGATCGCCGGCTGGCATGGGCGGGACCGGGT  
ACTTCGCCGGCGCCTCCGATATGTCCACCCCTGTTGACGACCTCGCCGCCGGCCACCGCCATGGC  
CTGGTGGCCCGCGTGTGCGAGCTGATACACCAGCGCTATCTGGCGTGTGAGGCCGACACTGATGCGGAGAC  
CGCGCGCGTCAACTGCGTACCGGGTACTCGGCCGTGGCTGCGAGGCCGATGTGCGGTAGCGCCGCC  
TCTCGTGGAGCTGCAGACCTCATGGAGTGGTTGCTCGGAATCGATATCCAGATCGGCTACGGATCCACC

**FIG. 8H**

GAGGCCGGTGGTGTATCCCGACGGAGTGGTCGTTGGCCGCCGGTCACGGAGTACAAGCTGATCGATGTCCCCGA  
 CCCGAACTGGGCTACTTCGTACCGACTCCCCCATCCACCGCGAAGCTCCTGGTCAGTCGACGCAGTGATTCCCGGGTACTACA  
 ACTCCGACAAGCGGATCCGACGACGAAGGCTCTACCGCACCGGATGTGACCAGATCTTCCCTGAGAACCAGCTCGCTGC  
 CGACCGAGCTGGGACCCGACCGGCTCGAGTACGTGACCGGGAGCAACGTGATCAAGTTGGCACAGGGAGAGTTCG  
 AGAGTTCGTGCAGTCGCCAACTCGAGGCCATCTACGCCGCCGGTCCCAGTGTGACCAGATCTTCCCTGAGAACCAGCTCG  
 CGACCGAGCTGGGACCCGCTACCGTACCGATGGCCTGGCGCATGCCGTGAGAACGATCTCGCTGCCTACGA  
 GGTGCCGCGCATGTGCTCATCGAACGTGATCCCTCTCAGGAGAACGGCTGGGTGGGGATCGGCA  
 AGCTGGTGCGCCCGGCCCTATCGCCCGTACGGTGACCGGTTGACGACCTCACGCCAGGCCACACC  
 CGTCAACCGAGGGCTTGCACGCGCTCGACGCCCTGGGCCATCATGACACCGTGCTCGGGCGGCTGC  
 GTTGACGCTCGGCCCGGATACTCGCGACTTCGACGCCGACACTCGATTCCGGGACCTCGGTGGGACTCGT  
 TGTCGGCGCTCGCTCGACGACGCTCGAAGGCCCTACGACGTGCCGTCCCGTGCAGACGATCGTC  
 GGACCGACGCCAACACTCGCGCGCTCGCCCGCACATCGAGAACGGCTCGATCGGTGGCGTC  
 GACCGCCGACTCGGTGCACGGCGTGGTGCAGCGTCGCCCGGGCACCGACCTGACGCTGGAGAACGTTCA  
 TCGACCCCGAGCTCTCGCGCTCGCCGACGCTCCCGCGGACCGGTGAGCGAACACCGTGCTGCTC  
 ACCGGATCCACCGGCTACCTCGGCCGTTCTCTGCTGCTGGACTGGTTGCGACGGGTGCTCCGACGGCG  
 CACCGTGAATCGCGCTGGTGCACGGCGACGCCGACGATGCGCAGCCGCTCACGGCGGATCGTG  
 ACTCGGATCCTGACCTGACACAAGAGTTCACTCGCGGAGCATCACCTCCACGTGATCGCCGGTGA  
 TTCGGCAGCCCGCACTCGGACTCGACGATGCCACTGGAGCGATCTCGCCGGGCGAGTCGATCACGTGG  
 GCACTCGGGCGCGCTCGTCAACCACGTGCTGCCCTACGACCAACTGTTGGTCCCAATGTTGGGCCACCG  
 GCGAAGTGGTGCAGTCGCACTCACCACGCGCCGCAAGTCCGTGGATTACGTCTCCACGGTGGCTGTGGTT  
 CGCAGGATGACGGCCGCTCTGGTCAAGGACGACGATGTTCGCGAGCTCGGCCGGAACGGCGCATCGG  
 GGCGATGCCTACCGAACGGCTACGCCGTGAGCAAATGGCGGGCGAAGTGTGTTGATGAGGCAGCCG  
 ACCTGGCGGACCTGCCGGTGCAGGTTCCGCTCCGATATGATCTGGCGCACAGTCGATTCCACGGACAG  
 TTCAACGAGGTGACCAAGTTCAACCGCCTGCTCTGAGTATCGCGAGACCGGACTGGCGCCGGCTCGT  
 CTACACGCCGGATCCGAGTGGACACCGCCGCACTACGACGGCTGCCGGTGGACTCACGCCGAAGCGA  
 TCACACGCTCAGCGCCGCGGGCGTCGGGGTACCGGACCTCCACGTGCTCAACGCCAACGATGACGGC  
 GTGAGCCTGGACAGCTCGTCACTGGATCGCCGCTCGGGCGGAGCATCGAACGGATCGACGACTACGA  
 CACCTGGTTGCCCGGTTCGAGCAGCGCTCCAGCAGCTCCCGATGAGGCAGCCGCGCCAGCGGTGCTGC  
 CCCTGCTGCACGCCGTGCGAGCCGCTCCGCCGGGACCTCCGCGCTGTCGGTGGACCGGTTCCG  
 GGTGCGGTGCGTGAAGACCGGAGTAGGACCGGGGACATCCGGTGCATCGGCCCTGATCGAGAACGTA  
 CCTGCGCACTCGAGACCGCGGCTGGCTCGGCCCGGTGCGCGCACTGA

**Amino acid sequence (SEQ ID NO:26)**

>gi|227980601|ref|ZP\_04027864.1| thioester reductase-like protein  
 [Tsukamurella paurometabola DSM 20162]

MSIETVQNGVPAEGSVPPADQQTERTLPQVIARIFAQFADRPAFATREAGPGTPYATVSYREIWRVRTALV  
 ASWQSEVAPGDFVAILGFTSSDFVTVDLATTLLGAPNVLQAGAPAARIATILDETRPKILAVSADQVDL  
 AQEALAEASAATPRVVVFDFGERDGYEGIEADILSGSALPAPEFFAPEPEPGTDPLVTLIYTSGSTGTPKGAMY  
 TEQLVRDAWLKVDSIVDIDMPAESLLHFLPMSHMYGRNWLIAGLASGGTYFAGASDMSTLFDDLAARP  
 TAIGIVPRVCELIHQRYLAVEADTDAETARVELRDRVLGGRQLQAAMCGSAALSSELQTFMEWLLGIDIQI  
 GYGSTEAGGVIRDGVVVRPPVTEYKLIDVPELGYFVTDSPHPRGEELVKSTQLIPGYYNSDKRIRDDEGF  
 YRTGDVMAELGPDRLEYVDRRSNVIKLAQGEFVPIAQLEAIYAAGPDVHQIFLYGTSERSYLIJVVVPAP  
 GPDGETDAQTRTRVLDGLAAIARENDAAYEVPRDVLIERDPFSQENGLRSGIGKLVRPALIARYGDRlh  
 DLYAQADTRQREGRLRALDASGPIIDTVLGAAALTIGADIADFADTRFGDLGGDSLSALSLATTLEGLYD  
 VPVPVQTIVGPTATLGGVARHIEKARSGGVAAPTADSVHVGASVARATDLTLEKFIDPELLALAPTLPA  
 ATGEPNVTLVLLGSTGYLGRFLLLDWLRVAPHGGTVIALVRGADADDARRVTAAIGDSDPDLTQEFTSL  
 AEHHHLVIAQDFGSPALGLDDATWSDLAGRVDHVVCALVNHLVYDQLFGPNVVATGEVVRLALTTRR

**FIG. 8I**

KSVDYVSTVAVVVPQDDGRVLVEDDVRELGAERRIGADAYANGYAVSKWAGEVLLHEAADLADLPVRVFR  
SDMILAHSRFHGQFNEVDQFTRLLSIAETGLAPASFYTPDPSGHRPHYDGLPVDFTAEAITLSAAGRS  
GYRTFHVLNANDGVSLDSFVDWIAASGRSIERIDDYDTWFARFEQALQQLPDEARQRSVLPILLHAVREP  
APAAGTSALSVDRFRGAVRETGVGPBDIPVLDRALIEKYLRDFETAGWLAPGARD

**ZP 05045132**

**Nucleotide sequence (SEQ ID NO:27)**

>gi|254430111:343253-346687 Cyanobium sp. PCC 7001 scf\_1106012173546  
genomic scaffold, whole genome shotgun sequence

GTGAATGAGTCTTCGGGACAGAGTTCCGGCAACGTTCCGAGGGGTGCCCTATGCTTCGGTCACAGC  
ACGGCCCTGCAGGCTCACCTGCCTACGAACAGATCATCGATGCCATTCTGAGCGGCTACGCCGAGCGC  
CTGCTCTGGCGAGCGTCTACCTGGTGCAGGCCGGACCGGAGCACAGGTCAAACGGTGCCTGTCACGAG  
CAGGCCCTCCGCTCGATCAGCTACCGAACCCCTGCAGGAACGGTTCATGCCCTCACCATGGCCTGGCGCCT  
TCATCCGATAGCCCGGTGCAAGCGGGAGCCTTCGTTGGTGGGATTGCCAGCATCGATTACGCCG  
TTCTTGATCTGGCACTGGCCTACACCAAGGGCGTGCCTGGTGCCTGTCACCGAACCATCCAGCGAGGAC  
GATGACGCCATCCCTCGGACAGTCCAGCCCCGTCACTCTGGCGTATCGATCAGTGGATCTCTGGCTGTG  
CGACTCTGATGCCCGATCGACGTCGATCCGAACCTGTGATCGCTTTGACCTTGACCTTGCCGTCGACTGCG  
AGCGCGCCGACTGGAGAGCGCATCCGGCACTCAACGAGAAGGGTCAGACGTTGCTCGTCAGACACTG  
CAGGATCTGATTGACGTTGGAGAGACGCAAGGTTCACTTCCTGCCGATCCAGCGCAGGATCAAGATGA  
CCTGGCACTTCTGATTCACACATCCGGCAGCACAGGCACACCCAAGGGAGCCTGCATCTCATCCCGTCAC  
TGATCAACACCTGGCGCCATGTTCCGGTCCCTATCCAAAAGTGACCGTGGTCTGGCACCCCTCCACCAC  
ATGATGGGACGAGACTCGATGATCACGGCATTGGGCGGGGGCACCCTCACTTCACGCTCAGGCCIGA  
CCTTCGACCGTGATTGAAGACATCAGACTGGCACGGCCCACAGGCCCTGGTGTGTTCCCGCCTCTGCG  
AAGTGATCGAACACCACCTGACTACTGGCGGGAGTATTCAAGGCAACGAGATCTCGGAGGGCAGACTGCAA  
TCCATTGTTGGTGGCCTCGGCTCCGATCACGCCACGCTTGAGGCATCCCTGGAGTGCCCTTGGGGTGC  
TGTGAGCGAAGGCTACAGCAGCACGGAAACAGCCAGTGGCGGGCTGGCGATGAATGGACTGCTGAACCGCA  
ACAACATTCTCGGTATGCCCTCGCGATGTGCCCAGGGCAGGGTATTCACTGAATGATCGGCCCTTCCG  
CGCGGAGAACTCTGCGTGAAGACCCGCTCGGTATCTCAGGCTATTCAGAAATCCAGAGGCCACTGCAGA  
GCTGTTGACGACGATGGCTTCTATTGACCGGTGACATCGTTGAAGAGCGGGCCCCGATCAAATCGCCA  
TCATCGACCGGGCAAAGAAATGTCATCAAGCTGGCGAGGGTGAATATGTCGCTGTGGCAGGCTGGAAACAG  
CTTTCCAGGAAGGTTGTGTTGCGTGAGCAGATTACCTCCACGGCGACAGCACAAGGGCTATCTGCT  
GGCAGTCGTTGTAACCTGATCGAACACCCCTTGACCGCCGGTACGGCAGGCCAGTGGAGGGCAGTTAA  
AGGCACGGGTGCGCGAGGGAGATTCTCACCTTGGCAAACCAACGGGAGCTGCGGGCTCGAGATCCCTCGA  
GACCTGATCTGGCGGAGGAACCCCTCAGCAGAACGGCTGCTGTGTCCTGGGTAAGCCGATCCG  
CCCGGCCATCCGGCGCGCTACCGCAGCCGGCTGGAGAGGCGCTATGCCAGCCATGAGGCCACCCGAGGA  
CTGAGCTCGAGGCCATCAGAGCGTCAGCTGGCGGGTGGATGTGAAACCCCTGTTGGCGCTGCTGAGC  
AGCACCGCTGGGTGTTGGTGTGGGCTGCCGATCGCAGACGAGTTCCGCGAGCTGGGGCAGTGGGGCGACTCC  
GGCGCTGTGCGACTGGAGATCAAGAAGCAGTTGGGGTGGGCTGGAGGGAGCCAGATTCTCG  
GGCGGGCGGACGGTGGAGCGTGGCGGGAGGATCCACACCGCCTCCATCCAGCAGGCCCGCACCAG  
CGGGTTGGCAGTCCCTCGGCCATTCCGGCCAGGGGTTGGCTGAAGCCGGACACTACAGGCTGGAGAA  
CCTGATCGGATCCCACCGTACACCTCAGCCAGGGTGGCAGGCCACAGGCGGGCCCTACGGTT  
TGCTCACCGGTGCCACCGTTCTGGGGGGCGCTTGTCGCTGGAGTGCTGCAACGGCTGGCTGGCCAG  
GGGGGCAGGCTGATCTGTCGGTGCCTCGAACAGCCATTCCGCTGGAGCGACTGAGGAACCGCTT  
CTCCCATCTGGAACCCGAGCAGGTGGCAGCTCCGCGAGCTGGCGGGAGGCTATGAGGTGATTCCGG  
CGGACATCGGAGAGGCCGGCTGGGCTTGAAACCGGTTGCCAGGAGCGGCTGCCACTGAGGTGGACGCG  
ATCTGTCACTGCGCAGCGGAAGTGAATCACCGGCTGCCCTATGCCACCTACCGGCCAATGTGATCGG  
CACCGCGGAGATCATTCACCTGGCGATCACGACGCCGTGAAATCGGTGGACTTCATCTCCAGCATTGGGG

**FIG. 8J**

TGGCTTCCCTGCCCGGCCGGAGGGAGCATTCCGGTGAGGGCGGCTACGCCCGGGCTACTTCGCC  
 AGCAAGTGGGCTTCGAGCAACTGCTGCGCTCCACCCATGACTGCACCGGTGTGCCGTACGGGTGATTG  
 GCCCAGCCTCATCTTCCCAGTCGTGCTGCCGGGGAGATGAACCCGGACGATCTGCTTCAGACTGC  
 TGTACAGCATCCTGGTGACCGGGATGCCCGGGGTGCTTGGGGAGGAGTCGCAGAACAGTGACGATCG  
 GGGTCTCGGTGCAGGGCCTCCCCGTGACCAGTGGCGCAGACCATCCTGCCCTCGGGAAAGCGCGCAC  
 GGAGGGATTTCATGTGCTAACCTCAACGCTGACAGTGGCAGCGGTGTTCCCCTGGATGCCATCCTCCAGG  
 ACATCGCCGCAAAGGAATCAGGCTGCGACGGGTTGAGGGCTATGACCTCTGGCTCGACGCGATCACACC  
 CGCCTCGCTCGCTGCCAGCCGAGCAACGGGCCCTCCCTGCTGGATGTCGGCGAAGCCTATGCAGGATC  
 AGCAGGCCAGACAACGCGAGCAGCGGTGAAATGCAGGCGGGCAGCAGCTCCTGCCCGGAGGAGATCACCA  
 GCCTGCAACCGGACTTCAGTAGGGCCTACAGGCCAAGATCGTGGATGATCTGGCTCGGTGGGGCTGATC  
 GAGCCTCCAGGACCCGTGGATCAGTGA

**Amino acid sequence (SEQ ID NO:28)**

>gi|254431429|ref|ZP\_05045132.1| putative long-chain fatty-acid--CoA  
 ligase [Cyanobium sp. PCC 7001]

MNESSADQSSGNVSEGWPDAVTARALQAHRLRYEQIIDAILSGYAERPALAERSYLVRPDPSTQTVRVH  
 EQAFRSISYRTLQERVHALTMARLHDSPVQAGAFVVLVGFASIDYAVLDLALAYTKGVPVPLSPNHSS  
 EDDDAILGTVQPVTLAVSISEFSGCVDLIARSTSIRTVIVFDLDPAVDCERAALLES GIRALNEKSDVV  
 QTLQDLIDVGRDAEFSFLPIQAQDQDDLALLIHTSGSTGPKGACISSRALINTWRHVSGPYPKVTVLA  
 PFHHMMGRDSMITALGAGGTAYFTLRPDLSVTIEDIRLARPTGLVLFPRLCVEVIEHLLTAPEYSNEIL  
 GGRLQSIVVASAPITPRLKASLECLLGVPVSEGYSTETASGLAMNGLNRNNILAYRLRDVPEAGYSV  
 NDRPFPRGEELCVKTRFGISGYFRNPEATAELFDDDGFYCTGDIVERAPDQIAIIDRRKNVIKLAQGEYV  
 AVGRLEQLFQEGCGCVQQIHLHGDSRAYLLAVVVDRNTLAPPGRSRQASEAELKARVREEILTLANQRE  
 LRGFEIPDLILAEPSQONGLSSILGKPIRPAIRARYRSRLESLYASHEATRGTELEAIRASAGADV  
 ETLLALLSSTLGVVCAGADRQTSFRELGGDSLAAVQLAMEIKKQFGVGLEGSQILGPGBTVEAWARRIH  
 TASIQQAPHQRVGSPLAAIPAEGWLKDHYRLENLIGIPIGTPSAEVARPTGGPPTVLLTGATGFLGGRL  
 CLEWLQLLAGQGGRLLICLVRPSNSHSAWERLRNRFSHLEPEQVARFRELAGRHLLEVIPADIGEPGLGLEP  
 GCQERLATEVDAICHCAAEVNHLRPLYRPNVIGTAEIILHAITTRLKSVDIFISSIGVASLPRRPGGS  
 IPVEGGYARGYFASKWACEQLLRSTHDCTGPVVRIRPSLILPDRVLAGEMPDDLSRLLYSILVTGIA  
 PGCFGEESQNNSGRSGFSVQGLPVDQLAQTILALGEARTEGFHVLNLNADSGSGVPLDAILQDIAAKGIRL  
 RRVEGYDIWLDAITTRLRLPAEQRARSLLDVAEAYAGSAGQTTSQSSGEMQAGSSSCPEEITSILQPDSR  
 AYRRKIVDDLARWGLIEPPGPVDQ

**YP 882653.1****Nucleotide sequence (SEQ ID NO:29)**

>uniprot|A0QIB5|A0QIB5\_MYCA1 Putative acyl-CoA dehydrogenase

ATGT'CGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACCTCATGCCACC  
 GACCCGCAATTGCCGCCGCCAACCGACCCGGCGATCACCGCCGCCCTCGAACAGCCC  
 GGGCTCGGGCTGCCGCAGATCATCCGACCGTGTGCGACGGCTACGCCGACCGGCCGGCG  
 CTGGGACAGCGCGTGGAGTTCGTACGGACCCAAGACCGGGCGCACGTCGGCGCAG  
 CTGCTCCCCCGCTTCGAGACCATTACGTACGGCAAGTGGCGCAGCGTGTTCGGCGCTG  
 GGCGCGCCCTGTCTGACGACGCGGTGCACCCGGCGACCGGGTGTGCGTGCTGGGCTTC

**FIG. 8K**

AACAGCGTCGACTACGCCACCATCGACATGGCGCTGGCGCCATCGCGCCGTCTCGGTG  
CCGCTGCAGACCAGCGCGCAATCAGCTCGCTGCAGCCGATCGTGGCCGAGACCGAGCCC  
ACCCCTGATCGCGTCCAGCGTGAACCAGCTGTCGACCGCGGTGCACTGATCACCGCGGCC  
GAGCAGGCGCCCACCCGGCTGGGGTTCGACTACCACCCGCAAGTCGACGACCAGCGC  
GAGGCCGTCAGGACGCCGCGCGCTGTCGGCACCAGCGTGGCCGTCAGACGCTG  
GCCGAGCTGCTGGAGCGCGCAAGGACCTGCCCGCCGTCGCGAGCCGCCGACGAG  
GACTCGCTGGCCCTGCTGATCTACACCTCCGGTCCACCGCGCCCCAAGGGCGCATG  
TACCCGAGAGCAACGTCGCAAGATGTGGCGCCGGCAGCAAGAACTGGTCGGCGAG  
AGCGCCGCGTCGATCACCTGAATTCTATGCCGATGAGCCACGTGATGGCCGAAGCATC  
CTCTACGGCACGCTGGCAACGGCGCACCGCTACTTCGCCGCCGAGCGACCTGTCC  
ACCCCTGCTCGAGGACCTCGAGCTGGTGCGGCCCACCGAGCTCAACTTCGTCGGGATC  
TGGGAGACGCTGTACCGCAATTCCAGCGTCAAGTCGAGCGGCCGCTCTCCGAGGCCGG  
GACGCCGCGAACGTCGCGCCGTCGAGGCCGAGGTGCTGGCGAGCAGCGCAGTACCTG  
CTGGCGGGCGGTTCACCTCGCGATGACGGGCTCGGCCCATCTGCCGAGCTGCGC  
AACTGGGTGAGTCGCTGCTCGAAATGCACTGATGGACGGCTACGGCTCACGGAGGCC  
GGAATGGTGTGTCACGGGAGATTCAAGGCCGCCGGTGGTCACTACAAGCTGGTC  
GACGTGCCGACCTGGCTACTTCAGCACCGGCCATCCGCGGGGAGCTGCTG  
CTGCGCACCGAGAACATGTTCCGGCTACTACAAGGGGCCAAACCACCGCGGGCGTC  
TTCGACGAGGACGGCTACTACCCACCGCGACGTGTTGCCAGAGATGCCCGGACCGG  
CTGGTCTACGTCGACCGCGCAACAACGTCGCTCAAGCTGGCGAGGGCAATTGTCACG  
CTGGCCAAGCTGGAGGCCTGTTCGGCAACAGCCGCTGATCCGCCAGATCTACGTC  
GGCAACAGCGCCACGGCTACCTGCTGGCGTGTGGCCACCGAGGAGGCCTGGCC  
TCGGGTGACCCCGAGACGCTCAAGCCAAGATGCCGACTCGCTGCCAGCAGGTC  
GAGGCCGGCCTGCACTACGAGGTGCCGCGACTTCATCATCGAGACCACCCGTC  
AGCCTGGAAAACGGCTGCTGACGGGATCCGAAGCTGGCGGCCAAACTGAAGCAG  
CACTACGGGAACGGCTGGAGCAGATGTAACGCCGACCTGGCCGCCAGGCCGACGAG  
CTGGCGAGCTGCGCCGCAACGGTGCCAGGCCGGTGTGCAAGACCGTGAGGCCGCC  
GCCGGCGCCATGCTGGGTTGCCGCCCTCGACCTGTCCCCGACGCCACTTCACCGAT  
CTGGCGGAGACTCGTTGTCGGCGTTGACATTGGCAACCTGCTGCCGAGATCTCGAC  
GTCGACGTGCCGGTGGCGTATCGTCAGCCCCGCCAACGACCTGGCGGCCATCGCAGC  
TACATCGAGGCCGAGCGCAGGGCAGCAAGCGCCGACGTTGCCCTCGGTGACGCCGG  
GACGCCGACCGTGGTGCAGCGCCGACCTGACGCTGGACAAGTCTCGACGCCGACACG  
CTGGCCTCCGCGCGAACCTGCCAAGCGGCCACCGAGGTGCGCACCGTGTGCTGACC  
GGGCCACCGGCTTCTGGGCCGCTACCTGGCCCTGGAAATGGCTGGAGCGGATGGACATG  
GTGGACGGCAAGGTATGCCCTGGTCCGGCCGCTCCGACGAGGAGGCACGCCGG  
CTGGACAAGACCTCGACAGCGGCACCCGAAGCTGCTCGCGACTACCAGCAGCTGGCT  
GCCGATCACCTGGAGGTATGCCCGCACAAGGGCAGGCCAATCTGGCCTGCCCAA  
GATGTTGCCAACGGCTGGCCGACACGGTCGACGTATCGTCGACGCCGCGCTGGTC  
AACCACGTGTTGCCGTACAGCGAGCTGTTGGGCCAACGCCCTGGCACCGCGAGCTG  
ATCCGGCTGGCGTACGTCAAGCAGAACGCCGTACACCTACGTTGTCACCGATGCCG  
GGCGACCGAGATCGAGGCCGGCAAGTTCGTCGAGAACGCCGACATGCCGAGATGAGGCC  
ACCCGGCGATCAACGACAGCTACGCCAACGGCTACGCCAACAGCAAGTGGCCGGCGAG  
GTGCTGCTGCCGAGGCCGACGCCGACTACGACGCCGCTGCCGTCAGGTTATGCCGCC  
ATGATCCTGGCGACACCGACGTATGCCGGCAGCTAACCTGCCGGACATGTTACCCGG  
CTGATGCTGAGCCTGGTGGCCACCGGGATCGCCGCCGGCTGTTACGAGCTCGACGCC  
GACGCCAACCGGCAAGGCCGACTACGCCAACAGCAAGCTGCCGCTGGTGGACGCC  
ATCTCGACCGCTGGGTTCGCAGATCACCGACAGCGACGCCGCTTCCAGACCTACCGTG  
ATGAACCCCTACGATGACGGCATCGGTCTGGACGAGTACGTCGATTGGCTGGACGCC  
GGCTATTGAGCGGATGCCGACTACCGAACGCCGAGTGGCTGCCGCTGCTGCCACA  
CTGCCGGCCCTGCCGGACCGGCCAGCGCCAGTACTCGCTGCTGCCGCTGCTGCCACA  
GGCACGCCGAGAACGCCGATCAACGGGTCGATAGCTCCCACCGACGTGTTCCGGGAGCG  
GTGCAGGAGGCCAAATGCCGGACAAAGACATTCCGACGCTGCTGCCGCCGGTATC  
GTCAAGTACATCACCGACCTGCAGCTGCTGGGCTGCTCTGA

**FIG. 8L****Amino acid sequence (SEQ ID NO:30)**

>uniprot|A0QIB5|A0QIB5\_MYCA1 Putative acyl-CoA dehydrogenase

MSTATHDERLDRRVHELIATDPQFAAAQPDPAITAALEQPGLRLPQIIRT  
VLDGYADRPALGQRVVEFVTDAKTGRRTSAQLLPRFETITYGEVAQRVSAL  
GRALSDDAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISS  
LQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTRLVVFDFYHPQVDDQR  
EAVQDAAARLSGTGVAVQTLLAELLERGKDLPAVAEPPADEDSLALLIYTS  
GSTGAPKGAMYPQSNGVKMWRGSKNWFGEESAASITLNFMPPMSHVMGRSI  
LYGTLGNGGTAYFAARSDLSTLLEDLELVRPTEILNFVPRIWETLYGEFQR  
QVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPISPELR  
NWVESLLEMHLMGDGYGSTEAGMVLFDGEIQRPPVVDYKLVDV рDLGYFST  
DRPHPRGELLRTENMFPGYYKRAETTAGVFDEDGYRTGDFVAFIAAPDR  
LVYVDRNNVNLKLAQGEFVTIАKLEAVFGNSPLIRQIYVGNSAQPYLLA  
VVVPTEALASGDPETLKPКIADSLQQVAKEAGLQSYEVPRDFIIETTPF  
SLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAGQADELAELRRNGAQ  
APVLTQTVSRAAGAMLGSAASDLSDAHTDLGGDSLSALTFCNLLREIFD  
VDVPVGIVISPANDLAAIASYIEAERQGSKRPTFASVHGRDATVVRAADL  
TLDKFLDADTLASAPNLPKPATERVTVLLTGATGFLGRYLALEWLERMDM  
VDGKVIАLVRARSDEEARARLDKFDSGDPKLLAHYQQLAADHLEVIAGD  
KGEANLGLRQDVWQRLADTVDVIVDPAALVNHVLPYSELFGNALGTAEL  
IRALALTSKQKPYTYVSTIGVGDQIEPGKVENADIROQMSATRAINDSYAN  
GYGNSKWAGEVLLREAHDLCLGLPVAVFRCDMILADTTYAGQLNLPMFTR  
LMLSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAISTLGSQITD  
SDTGFQTYHVMNPYDDGIGLDEYVDWLVDAGYSIERIADYSEWLRRFETS  
LRALPDRQRQYSLLPLLHNYRTPEKPINGSIAPTDVFRAAVQEAKIGPDK  
DIPHVSPPVIVKYITDLQLLGLL

**YP 887275.1****Nucleotide sequence (SEQ ID NO:31)**

>uniprot|A0QWI7|A0QWI7\_MYCS2 NAD dependent epimerase/dehydratase family protein

ATGACGATCGAAACGCGCGAAGACCGCTTCACCCGGCGCATGTACCAC TTGTTCGAAACC  
GACCCGCAGTTGCCGCCGGTCAAACAGATTCTGGCCGGCTATCGGGACCGCCCTGCG  
GAGTTGCCCTTCTGCCGCCGGTCAAACAGATTCTGGCCGGCTATCGGGACCGCCCTGCG  
CTGGGCAAGCGGCCGTCGAGTTCGTACCGACGAAGAAGGGCGCACCCCGGAAGCTC  
CTGCCCCGCTTCGACACCACACCTACCGTCAGCTCGCAGGCCGGATCCAGGCCGTGACC  
AATGCCCTGGCACACCACCGTGAATGCCGTGACCGCGTGGCATCCTGGTTTCACC  
AGTGTGACTACACGACGATCGACATGCCCTGCTCGAАCTCGCGCCGTGTCCGTACCG  
CTGCAGACCAAGTGCGCCGGTGGCCAАCTGCAGCCGATCGTCGCCAGACCGAGCCCAAG  
GTGATCGCGTCGAGCGTCGACTTCCTCGCCACGCGACTCGCTCTCGTCGAGTCCGGCCC  
GCGCCGTGCGACTGGTGGTTCGACTACAGCCACGAGGTGACGATCAGCGTGAGGCG  
TTCGAGGCCAAAGGGCAAGCTCGCAGGCACCGCGTCGTCGAGACGATCACCGAC  
GCACTGGACCGCGGGCGGTCACTCGCCACGCAACCGCTCTACGTGCCGACGAGGCCGAC  
CCGCTGACCCCTCTCATCTACACCTCCGGCAGCACCGCACTCCCAAGGGCGGATGTAC  
CCCGAGTCCAAGACCGCCACGATGTGGCAGGCCGGTCCAAGGCCGGTGGGACGAGACC

**FIG. 8M**

CTCGGC GTGATGCCGTGATCACCC TGAACTTCATGCCCATGAGTCACGT CATGGGGCGC  
GGCATCCTGTG CAGCACACTCGCCAGCGCGGAACCGCGTACTCGCCGCACGCAGCGAC  
CTG TCCACCTTCTGGAGGACCTCGCCCTCGTGGGCCCACCGCAGCTCAACTCGTCCT  
CGCATCTGGGACATGCTGTTCCAGGAGTACCA GAGCCGCTCGACAACC CGCGCGAG  
GGATCCGAGGACCGAGCGAAGCCGAGTCCTCGAAGAGGTCGCACCCA ACTGCTCGC  
GGCGATTGTT CGGCCCTGACCGGATCGGCTCCCATCTGGCGGAGATGAAGAGCTGG  
GTCGAGGACCTGCTGACATGCATCTGCTGGAGGGCTACGGCTCCACCGAGGCCGCG  
GTGTTCATCGACGGCAGATCCAGCGCCGCCGGTACATCGACTACAAGCTGGTCACGTG  
CCCGATCTGGCTACTTCGCCACGGACCGCCCTACCGCGCCGAACTTCTGGTCAAG  
TCCGAGCAGATGTTCCCGCTACTACAAGCGTCGGAGATCACCGCCGAGATGTTGAC  
GAGGACGGGTACTACCGCACCGCGACATCGTCGGAGCTGGGCCGACCATCTCGAA  
TACCTCGACCGCCGCAACACGTGCTGAAACTGTCGAGGGCAATTCTGTCACGGTCTCC  
AAGCTGGAGGC GGTT CGGCACAGCCCCCTGGTACGCCAGATCTACGTCTACGGCAAC  
AGCGCGCGGT CCTATCTGCTGGCGGTCTGGTCCCGACCGAAGAGGCACTGTCACGTTGG  
GACGGTGACGA ACTCAAGTCGCGCATCAGCGACTCACTGCAGGACGCCACGCC  
GGATTG CAGTCGATGAGATCCCGCGT GACTTCCCTCGTCA GAGACAACACCTT CACGCTG  
GAGAACGGC CTGCTGACCGGTATCCGCAAGCTGGCCCGGCGAAACTGAAGGCGCACTAC  
GGCGAACGCCTCGAACAGCTCTACACC GACCTGGCGAGGGCAGGCCAACGAGTTGCGC  
GAGTTGCGCCGCAACGGAGCGACCGGCCGTGGTCA GAGACCGTCA GCGCGCCGCC  
GCACTGCTGGTGCCTCCGTACGGATCTGCGGTCGATGCGCACTTCACCGATCTGGT  
GGAGATTGCTGTCGGCTTGAGCTCTCGAACCTGTTGACGAGATCTCGATGTCGAC  
GTGCCGGT CGGCGTCA TCGTCA GCGCCGCC ACCGACCTGGCAGGC TCGCCGCTACATC  
GAGGGCGA CTGCGCGCTCAAGCGCCCCACATACGCGTCGGTGCACGGCGCGACGCC  
ACCGAGGTGCGCGCGCGT GATCTGCCCTGGCAAGTTCATCGACGCCAACCGCTGTCC  
GCCCGCGCGGGTCTGCCGCTTCGGCACCGAGATCCGCACCGT GCTGACCGCGCC  
ACCGGGTCTGGCGCTATCTGGCGCTGGAAATGGCTGGAGCGCATGGACCTGGTGGAC  
GGCAAGGTGATCTGCCCTGGTGC CGCCGCCAGCGACGACGAGGCCCGGGCGCTGGAC  
GCCACGTTGACACCGGGGACGCGACACTGCTCGAGCACTACCGCGCGTGGCAGCCGAT  
CACCTCGAGGTGATCGCCGGT GACAAGGGCGAGGGCGATCTGGTCTCGACCGACACG  
TGGCAGCGACTGCCGACACCGTGATCTGATCGTCACTGGCCGCCCTGGTCAATCAC  
GTCCTGCCGTACAGCCAGATGTTCGGACCCAATCGCGTCGGCACCGCGAACACTCATCGG  
ATCGCGCTGACCA CCGATCAAGCCGTACGTGATCGACGATCGTGTGGACAG  
GGCATCTCCCCGAGGC GTGCTCGAGGACGCCACATCCCGAGATCAGCGCGACCGC  
CGGGTCGACGACTCGTACGCCAACGGCTACGGCAACAGCAAGTGGCCGGAGGTCTG  
CTGCCGGAGGCCACGACTGGTGTGGTCTGCCGGTCTCGGTTCTACGAACCTCGATGCGGACGGC  
AACCGGCAGCGCGCCCACTACGACGGCTGCCGTGGAGTTATCGCCGAGGCATCTCC  
ACCATCGGCTCGCAGGT CACCGACGGATTGAGACCGTCACTCGATGAGAACCGTACGAC  
GACGGCATCGGCCCTCGACGAGTACGTGGACTGGCTGATCGAGGCCGGCTACCCGTGCAC  
CGCGTCGACGACTACGCCACCTGGCTGAGCGGTCTGAAACCGCACTGCGGGCCCTGCCG  
GAACGGCAACGT CAGGCCCTCGTCTGCCGTGCTGCACA ACTATCAGCAGCCCTCACCG  
CCCGTGTGCCGGTCCATGGCACCCACCGACCGGTTCOGTGCCGCGGTGCAGGACGCCAAG  
ATCGGCCCCGACAAGGACATTCCGCACGTCA CGGCCGACGTGATCGTCAAGTACATCAGC  
AACCTGCAGATGCTCGGATTGCTGTA A

**Amino acid sequence (SEQ ID NO:32)**

>uniprot|A0QWI7|A0QWI7\_MYCS2 NAD dependent epimerase/dehydratase family protein

MTIETREDRFNRRIDHLFETDPQFAAARPDEAISAAAADPELRLPAAVKQ

FIG. 8N

ILAGYADRPALGKRAVEFVTDEEGRTTAKLLPRFDITIYRQLAGRIQAVT  
NAWHNHPVNAGDRVAILGFTSDYTTIDIALLELGAVSVPLQTSAVPAQL  
QPIVAETEPKVIASSVDLADAVALVESGPAPSRLVVFDSHEVDDQREA  
FEAAKGKLAGTVVVETITDALARGRSLADAPLYVPDEADPLTLIYTSG  
STGTPKGAMYPESKTATMWQAGSKARWDETGLVMPSTITLNFMPSHVMGR  
GILCSTLASGGTAYFAARSDLSTFLEDLALVRPTQLNFVPRIWMLFQEY  
QSRLDNRRRAEGSEDRAEAAVLEEVRTQLLGGRFVSAUTGSAPISAEMKSW  
VEDLLDMHLLEGYGSTEAGAVFIDGQIQRPPVIDYKLVDVPDLGYATDR  
PYPRGELVKSEQMFPGYYKRPEITAEMFDEDGYYRTGDIVAELGPDHLE  
YLDRRNNVLKLSQGEFTVSKLEAVFGDSPLVRQIYVGNSARSYLLAVV  
VPTEEALSRWDGDELKSRISSDLQDAARAAGLQSYEIPRDFLVETTPFTL  
ENGLLTGIRKLARPKLKAHYGERLEQLYTDLAEQGQANEIRELRRNGADRP  
VVETVSRAAVALLGASVTDLRSDAHFTDLGGDSLSALSFSNLLHEIFDVD  
VPVGIVVSPATDLAGVAAYIEGELRGSKRPTYASVHGRDATEVRARDLAL  
GKFIDAKTLSAAPGLPRSGTEIRTVLLTGATGFLGRYLALEWLERMDLV  
GKVICLVRARSDDEARARLDATFDTGDATLLEHYRALAADHLEVIAGDKG  
EADLGLDHDHTWQRЛАDTVДLIVDPАALVNHVLPYSQMFGPNALGTAELIR  
IAЛTTIКPVYVYSTIGVGQGISPEAFVEDADIREISATRRVDDSYANGY  
GNSKWAGEVLLREAHWCGLPVSVFRCDMILADTTYSGQLNLPDMFTRLM  
LSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAEAISTIGSQVTDF  
ETFHVMNPYDDGIGLDEYDVWLIЕAGYPVHRVDDYATWLSRFETALRALP  
ERQRQASLLPLLHNYQQPSPPVCAGAMPTDRFRAAVQDAKIGPDKDIPH  
TADVIVKYISNLQMLGLL

ZP 05224908

### Nucleotide sequence (SEQ ID NO:33)

>gi|163719654:2489-6013 Mycobacterium intracellulare ATCC 13950  
NZ ABIN01000072, whole genome shotgun sequence

ATATGTCGACTGCCATTCTATGACGAACACCTCGACCGTCGCATCGAGGAACCTCATCGCCAACGACCCCCAAAT  
TCGCGCCGCCGGACCGGACCCGGCCATACCGCCGCCACCGAAGCGCCGGGCTGCGGCTGCCAGAT  
CATCCGGACCGTGTGACGGCTACGCCGACCGGCCCTGCCCTGGCGCAGCGCTGTGGAGTTCTGCACC  
GACGCCAAGACCGGGCGGACGGCCGAGCTGCTCCCCGTTCTGAGACCATCACGTATGGCGAAGTCG  
GCGAACGGGTTTCGCCCTGCCGTGCCCTGGGCCGACGGGTGCGCCCCGGGACCGCGTCTGCGT  
GCTCGGCTTCAACAGCGTTGACTACGCCACCATCGACATCGCGTGGCACCATCGGGGCCGTGCGGTG  
CCGCTGCAGACCAGCGGGCGATCTCTCGTTGCAGCGATCGTGCCTGAGACCGAGGCCAGCCTGATCG  
CCTCGAGCGTCAACCAGCTGCCCGACCGGGTGAGCTGATCCTGCCCGGACACGTGCCCGGAAAGCT  
CGTCGTGTTGACTACAGCCCCAGGTGACGACCGACGGCGAGGCCGCGCCGGTTG  
GCCGACTCCGGCGTGCAGGTCTCGCCGACGTGCTGCCGCGGCAAGGACCTGCCGGCGTGC  
AGCCGCCGGCGAGCGACGAGGACTCGCTGCCCTGCTGATCTACACCTCCGGCAGCACCGCGGCCCAA  
GGCGCGATGTACCGCAGAGCAACGTGGCAAGATGTGGCGGGCGGGGAGCAAGAACCTGGTTGCCGGAA  
AGCGCCCGTGTGATCACCTCAACTCATGCCATGAGCCACGTGATGGGCCGGAATCCTCTACGGCA  
CGCTGGCAACGGCGCACCGCGTACTTCGCCGCCGAGCGACCTGTCACCCCTGCTGAGGACCTCGA  
GTTGGTGCAGGCCACCGAGATGAACTTCGTCCCCCGATCTGGAGACGCTGTACGGCGAATTCCAGCGC  
CAGGTCGAGCGCGGCTGGCCACGGCGATGCCGGCCGGAGGCGAGACTGTGGCGGCTGCCGTG  
TGGAAAGAACAGCGCCAGTACCTGCTGGCGGGCGGTCACTCTCGCGATGACGGCTCGCACCCACCTC  
GCCGGAGCTCAAGGGCTGGGCCGAGTCGCTCCTGCAGATGACCTGATGGACGGCTACGGCTCCACCGAG  
GCCGGAATGGTGTGTTGACAGGGGAGATTAGCGTCCGCCGGTATTGATTACAAGCTGGTGCACGTT  
CGGATCTGGCTATTCAGCACCGACCGTCCGATCCGCGCGGTGAGTTGTTGCTGCCGGACCGAGAACAT  
GTTCCCGGGTATTACAAGCGGGCGAGACCAAGCGAACGTGTTGACGAGGAGCGGTATTACCGCACC

**FIG. 80**

GGTGACGTGTTCCGCCAGAGATCGCGCCGGACCGGCTGGTGTATTCGATCGCCGAAACAACGTGCTCAAGT  
TGGCCCAGGGCAGGTCGTGACGCTGGCAAGCTGGAGGCGGTGTCGGCAACAGCCGCTGATCCGCCA  
GATCTACGTTACGCCAACAGCTCCAGCCCTACCTGCTGGCGTGGTGGTGCACCGAGGAAGCGTTG  
GCGGACAACGATCTGAGTCGCTCAAGCGAAGATCGCGACTCGCTGAGAAGTCGCCAAGGGAGACCG  
GCCTGCGACTCCTACGGGTGCCCGCAGCTCATCATCGAGACACGCCGTTCACCCCTGGAAAACGGCCT  
GCTGACCGGGATCCGCAAGCTGGCGTGGCCAAGCTCAAGGCCACTACGGGATCGCTGAGCAGATG  
TATGCCGAGCTGGCCCGGGACAGGCCAACGAGTTGGCGAAGTCGCCAGCGGGCGGGCGCCGG  
TGGCCCAAGACCGTGAGCCGGCCGCCGCCCTGCTGGTGCACGCCGGGATCTGTCGCCAGATGC  
CCACTTCACCGATCTGGTGGAGACTCGTTGTCGGCGTTGACCTTCGGCAACCTGTCGCCAGATCTTC  
GATGTCGACGTCGCCGGTGGGGGTGATCGTCAGCCCCGCCAACGACCTGGCGGGGATGCCGCTACATCG  
AGGCCGAGCGGCAGGGCTCCAAGCGCCCGACGTTGCCCGCGTGCACGGTGCCTGCGACCATGGTGA  
CGCCAGTGAACCTCACGCTGGACAAGTTCTCGACGAGGCAGCCCTGGCCGCCGCCAGCTGCCAAG  
CCGGCCACCGAGGGTGCACCGTGTGTTGACCGGGCGACCCGGCTTTTGGGCCGTAACCTGGCGCTGG  
ACTGGCTCGAGCGGATGGACATGGTCGACGGCAAGGTCACTGCCCTGGTGCAGGGCCCGACCGATGAGGA  
GGCGCGCCGGCTGGACAAGACCTCGACAGCGCGACCCAACTGCTGGCGACTACCAGCGGCTG  
GCCGCCGACCACCTCGAGGTCACTGCCGGGACAAGGGTAGGGCAACCTGGCCTGGACCCCCAGACCT  
GGCAGCGACTGGCCGAGGAGGTGACGTATCGTCGACCCCGCCGCGTGGTCAACCACGTGCTGCCCTA  
CAGCGAGCTGTTGCCACCGTGGCACCGGGAGCTGATCCGGATCGCGCTGACCTCCAGGCAA  
AAGCCCTACACCTACGTGTCGACGATCGGGTGGCGATCAGATCCAGCCAGGTGAGTTCTGTCGAGAACG  
CCGACATCCGCCAGATCAGGCCACCGCGAGATCAACGACGGTACGCCAACGGCTACGCCAACAGCAA  
GTGGGCCGGCGAGGTGTTGTCGCGAGGCCACGACCTGTCGGGCTGCCGTCACGGTGTCCGCTG  
GACATGATCCTGCCACCCACTATGCCGGCAGCTAACCTGCCGACATGTTCACCCGGCTGATG  
TGAGCCTGGTCGCCACCGGATCGCGCCGGGTCGTTCTACGAACGGCCGACGCCAACGCCAGCG  
GGCACACTACGACGGTTGCCGTCGAGTTCATGCCCGCGATCTGACGCTGGGACCCAAATCACC  
GACAGCGACACGGGTTTCAGACCTACGACGATGAAACCCCTACGACGACGGCATGGCTGGATGAGT  
ACATCGATTGGCTGATCGAGGGGGTATTCGATCGAGCGGATGCCGATTACTCGAGTGGCTGCCG  
CTTCGAGACCTCGCTGCCGGCCCTGCCGATCGGCGACGCTGAGTACTCGCTGCCGCTGCTGACAAC  
TACCAAGAGCCGGAAAAGCCGATCAACGGCTCGATGGCGCCACCGACGTTCCGTTGCCGCGGTGCAGG  
AAGCGAAAATCGGCCCGACAAAGACATCCCGCACGTCGCGCCGGTGATCGTCAGTACATCACC  
CCTGGAGTTGCTGCCGACTCCTCTGA

### Amino acid sequence (SEQ ID NO:34)

>gi|254819907|ref|ZP\_05224908.1| FadD9 [Mycobacterium intracellulare ATCC 13950]

MSTAIHDEHLDERRIEELIANDPQFAAARPDPAITAATEAPGLRLPQIIRTVLDGYADRPALAQRVVFVTD  
DAKTGRTTAELLPRFETITYGELGERVSALGRAWAGDAVRPGDRCVLGFNSVDYATIDIALGTIGAVSV  
PLQTSAAISSLQPIVAETEPSLIASSVNQLPAVELILAGDHVPKGKLVVFDYQPQVDDQREAVEAAAARL  
ADSGVAVEALADVLRRGKDLPAVEPPASDEDSLALLIYTSGSTGAPKGAMYPOSNVGKMWRRGSKNWFGE  
SAASITLNFMPPMSHVMGRGILYGTGLNGGTAYFAARSDLSTLLEDLELVRPTEMNFVPRIWETLYGEFQR  
QVERRLAGDGAGPEARETVAAVLEEQRQYLLGGRFIFAMTGSAPTSPELKAWAESLLQMHLMGDGYGSTE  
AGMVLFDGEIQRPPVIDYKLVDVPDLGYFSTDPRPHRGELLRTENMFPGYYKRAETTANVFDEDGYYRT  
GDVFAEIAPDRLVYDRRNNVLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSSQPYLLAVVVPTEEAL  
ADNDLESLKPCKIADSLQKVAKETGLQSYPEVPRDFIETTPFTLENGLLTGIRKLAWPKLKAHYGDRLEQM  
YAELAAGQANELAELRRSGAACAPVAQTVSRAAAALLGATAGDLSADAHFTDLGGDSLSALTFGNLLREIF  
DVDVPVGIVVSPANDLAGIAAYIEAERQGSKRPTFAAVHGRGATMVHASDLTLDFKLDTEATLAAAPSLPK  
PATEVRTVLLTGATGFLGRYLALDWLERMDMVDGKVIALVRARTDEEARARLDKTFDSGDPKLLAHYQRL  
AADHLEVIAKGDKGEANLGDPQTWQLAEEVDVIVDPAALVNHVLPYSELFGPNALGTAELTRIALTSRQ  
KPYTYVSTIGVGDQIOPGEFVENADIROQISATREINDGYANGYGNSKWAGEVLLREAHDCGLPVTVFRC  
DMILADTTYAGOLNLPMDFTRLMLSLVATGIAPGSFYELDADGNRORAHYDGLPVEFIAAAISTLGTGTOIT

**FIG. 8P**

DSDTGFQTYHVMNPYDDGIGLDEYIDWLIEAGYSIERIADYSEWLRRFETSLRALPDRQRQYSLLPLLN  
YQKPEKPINGSMAPTDVFRAAVQEAKIGPDKDIPHVSAPVIVKYITDLELLGLL

**YP 889972.1**

**Nucleotide sequence (SEQ ID NO:35)**

>uniprot|A0R484|A0R484\_MYCS2 Putative long-chain fatty-acid--CoA ligase

ATGACCAGCGATTTCACGACGCCACAGACGGCGTCACCGAAACCGCACTCGACGACGAG  
CAGTCGACCCGCCGCATCGCCGAGCTGTACGCCACCGATCCCGAGTCGCCGCCGCCGCA  
CCGTTGCCCGCCGTGGTCACGCGGGCACAAACCCGGGCTGCCGCTGGCAGAGATCCTG  
CAGACCTGTTCACCGGCTACGGTACCGCACCCTGGCGCTGGGATACCGCACCCTGA  
GCCACCGACGAGGGCGGGCGCACCGTGACCGTCTGCTGCCGCGTTGCCGACACCCCTACC  
TACGCCAGGGTGTGGTCGCCGTGCAAGCGGTGCCGCGCCCTGCCGACAAACTCGCG  
CAGCGATCTACCCCGCGACGCCGTGCCGACGATCGGTTGCCGAGTCCCATTACCTG  
ACGCTGGATCTCGTATGCCCTACCTGGGCTCGTGAGTGTCCGCTGCCGACACA  
CCGGTCAGCCGGCTGCCCGATCTGGCGAGGTGCAACCCGGGATCCTCACCGTGAGC  
GCCGAATACTCGACCTCGCAGTCGAATCCGTGCCGGACGTCACACTCGGTGTCGAGCTC  
GTGGTGGTCGACCATCACCCGAGGTGACGACCAACCGCAGCAGCAGTGGCCGCGCGT  
GAACAACCTGCCGGCAAGGGCATGCCGTACCACCCCTGGACCGGATGCCGACGAGGGC  
GCCGGCTGCCGCCGAACCGATCTACACGCCGACCATGATCAGCGCTCGCGATGATC  
CTGTACACCTCGGGTCCACCGCGCACCCAAGGGTGCAGTGTACACCGAGGCGATGGT  
GCCGCGCTGTGGACCATGTCGTTCATACGGGTGACCCACGCCGGTCAACGTCAAC  
TTCATGCCGCTCAACCACCTGGCGGGCGATCCCCATTCCACCGCCGTGAGAACGGT  
GGAACCACTTACCGTACCGGAATCCGACATGTCACGCTGGTGAGGATCTCGCGCTG  
GTGCGCCCGACCGAACTCGGCTGGTCCCGCGCTGCCGACATGCTCTACCAGCACCAC  
CTCGCACCGTGCACCGCCTGGTCACCGCAGGGCGCCGACGAACTGACCCGGAGAACAG  
GCCGGTGGCAACTCGGTGAGCAGGTGCTCGGCGACCGTGATCACCGGATTCGTCAGC  
ACCGCACCGCTGCCGCGGAGATGAGGGCGTCTCGACATCACCTGGCGCACACATC  
GTCGACGGCTACGGGCTACCGAGACCGCGCCGTGACACCGCAGGGTGTGATCGCG  
CCACCGGTGATCGACTACAAGCTGATCGACGTTCCGAACTCGCTACTTCAGCACCGAC  
AAGCCCTACCCCGTGGCGAACTGCTGGTCAGGTGCGCAAACCGCTGACTCCGGGTACTAC  
AAGCGCCCCGAGGTACCGCGAGCGTCTCGACCCGGACGGCTACTACCAACACCGCGAC  
GTCATGGCGAGACCGCACCCGACCTGGTGTACGGACCGCTCGCAACAGTCCTC  
AAACTCGOGCAGGGCGAGTTGTCGTTGGCGCTGCCAACCTGGAGGCGGTGTTCTCGCG  
GCGCTGGTGGCGCAGATCTCGTGTACGGCAACAGCGAGCGCAGTTCTCTGGCGTG  
GTGGTCCCGACGCCGGAGGCCTCGAGCAGTACGATCCGGCGCGCTCAAGGCCGCGCTG  
GCCGACTCGCTGCCGACCGCACCGCAGCGAACCTGCAATCCTACCGAGGTGCCGGCC  
GATTTCATCGTCAAGACCGAGCCGTTAGCGCCGCCAACGGCTGCTGCGGTGTCGGA  
AAACTGCTGCCGCCAACCTCAAAGACCGCTACGGGCGCCCTGGAGCAGATGTACGCC  
GATATCGCGGCCACGCAGGCAACCAGTTCGCGCAACTCGCGCGCGGCCACACAA  
CCGGTGTGACACCCCTACCCAGGCCGCTGCCACGATCCTCGGCCACGGGAGCGAGGTG  
GCATCCGACGCCAACCTCACCGACCTGGCGGGGATTCCCTGTCGGCGCTGACACTTCG  
AACCTGCTGAGCGATTCTCGGTTGCAAGTCCGTCGGCACCATCGTAACCCGGCC  
ACCAACCTGCCAACCTCGCCACGACATCGAGGCGCAGCGCACCGCGGGTGACCGCAGG  
CCGAGTTTACCAACCGTGCACGGCGGGACGCCAACCGAGATCCGGCGAGTGAGCTGACC  
CTGGACAAGTTCATCGACGCCAACCGCTCCGGGCCAACCGGGTCTGCCAACGGTCA  
ACCGAGGCCACGGACGGTGTGCTCTGGCGCCAACGGCTGGCTGGCGGTTCCCTCACG  
TTGCACTGGCTGGAACGCCCTGGCACCTGTCGGCGCACCCCTCATCACGATCGTGC  
CGCAGACGCCGGCGGCCACGGCTGACCCAGGCCTACGACACCGATCCGAGTTG  
TCCCGCCGCTTCGCCAGCTGGCGACCGCCACCTGCCGGTGGTCGCCGGTGACATCGG

**FIG. 8Q**

GACCCGAATCTGGGCCTCACACCCGAGATCTGGCACCGGCTCGCCGCCAGGTCGACCTG  
GTGGTGCATCCGGCAGCGCTGGTCAACCACGTGCTCCCCTACCGCAGCTGTTGGCCCC  
AACGTCGTGGCAGGGCGAGGTGATCAAGCTGGCCCTACCGAACGGATCAAGCCCCTC  
ACGTACCTGTCCACCGTGTGGCATGGGATCCCCGACTTCGAGGAGGACGGCAG  
ATCCGGACCGTGAGCCGGTGCAGCCCGCTCGACGGCGGATACGCCAACGGCTACGGCAAC  
AGCAAGTGGGCCGGCGAGGTGCTGCTGGGGAGGCCACGATCTGTGCGGGCTGCCGTG  
GCGACGTTCCGCTCGGACATGATCCTGGCGCATCCGCGTACCGCGTCAGGTCAACGTG  
CCAGACATGTTACCGCGACTCCTGTTGAGCCTCTGATCACCGCGTCCGCCGGTCG  
TTCTACATCGGAGACGGTGAGCGCCCGGGCGCACTACCCGGCCTGACGGTCGATTTC  
GTGGCCGAGGCGGTACGACGCTGGCGCGCAGCAGCGCAGGGATACTGTGTCCTACGAC  
GTGATGAACCCGACGACGGATCTCCCTGGATGTGTTGACTGGCTGATCCGG  
GCAGGCCATCCGATCGACGGGTCGACGACTACGACGACTGGGTGCGTCGGTTGAGACC  
GCGTTGACCGCGCTCCCGAGAAGCGCCGCGCACAGACCGTACTGCCGCTGTCACGCG  
TTCCCGCCTCCGAGGCACCGTGCAGGGCAGCCGAACCCACGGAGGTGTTCCACGCC  
GCGGTGCGCACCGCGAAGGTGGGCCGGAGACATCCGACCTCGACGAGGCGCTGATC  
GACAAGTACATACCGGATCTCGTGAGTCGGTCTGATCTGA

**Amino acid sequence (SEQ ID NO:36)**

>uniprot|A0R484|A0R484\_MYCS2 Putative long-chain fatty-acid--CoA ligase

MTSDVHDATDGVTETALDDEQSTRRIAELEYATDPEFAAAAPLPAVVDAAH  
KPGLRLAEILQTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDLT  
YAQVWSRVQAVAAALRHNFQAQPIYPGDAVATIGFASPDYLTLSDLVCAYLG  
LVSVPLQHNAPVSRLAPILAEEVEPRIITVSAEYLDLAVESVRDVNSVSQL  
VVFDHHPGVDDHHDALARAREQLAGKGIATVTLDAIADEGAGLPAEPIYT  
ADHDQRLAMILYTSGSTGAPKGAMYTEAMVARLWTMSFITGDPTPVINVN  
FMPLNLHGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVP  
RVADMLYQHHLATVDRLVTQGADELTAEKQAGAEIREQVLGGRVITGFVS  
TAPLAAEMRAFLDITLGAHIVDGYGLTETGAVTRDGVIVRPVIDYKLID  
VPELGYFSTDKPYPRGEELLVRSQTLTPGYYKRPEVTASVFDRDGYHTGD  
VMAETAPDHLVYDRRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYG  
NSERSFLLAVVVPTPEALEQYDPAALKAAALADSLQRTARDAELOQSYEVPA  
DFIVETEPFSAANGLLSGVGKLLRPNIKDRYQQRLEQMYADIAATQANQL  
RELRRRAAATQPVIDTLTQAAATILGTGSEVASDAHFTDLGGDSLSALTLS  
NLLSDFFGFEVPVGTIVNPATNLAQLAQHIEAQRTAGDRRPSFTTVHGAD  
ATEIRASELTLDKFIDAETLRAAPGLPKVTTEPRTVLLSGANGWLGRFLT  
LQWLERLAPVGGTLITIVGRRDDAARARLTQAYDTDPELSRRFAELADR  
HLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGP  
NVVGTAEVIKLALTERIKPVTYLSTSVAMGIPDFEEDDIRTVSPVRLP  
DGGYANGYGNISKWAGEVLLREAHDLCLGPVATFRSDMILAHPRYRGQNV  
PDMFTRLLSLLITGVAPRSFYIGDGERPRAHYPLTVDFVAEAVTTLGA  
QQREGYVSYDVMNPHDDGISLDVFVDWLIRAGHPIDRVDDYDDWVRRFET  
ALTALPEKRRRAQTVLPLLHAFRAPQAPLRGAPEPTEVFHAAVRTAKVPGP  
DIPHILDEALIDKYIRDRLREFGLI

FIG. 8R

ZP 04751860

### Nucleotide sequence (SEQ ID NO:37)

>gi|218125542:1370-4894 Mycobacterium kansasii ATCC 12478  
NZ ACBV01000156, whole genome shotgun sequence

**FIG. 8S**

```

gttctacgagctggacgcccacggcaaccggcagcgctgcccactacgacggcttgcggcgtcagttcatcg
ccgaggcgatcgcacgcgtggcgccccggacggaaagggttccagacactaccacgtatgaaccttac
gacgacggcatcggtatggaccgggttcgtcactggctcgtcagcgcggatgcgcattccaccgcac
cgactacggcactggctgcacgattcggagaccgcgtgcggcgtcccggaaaagcagcgtcacgcgt
cactactccgttgctgcacaactaccagaaggccggccgcgtgcgggtcgatggctccgaccgac
cggttccggccggccgtgcaggacgcgaaagtggccggacaaggacatccgcacatctcgccgcagat
catcgcaagtacctcagtatctgcgttcgtgcgttcctctga

```

**Amino acid sequence (SEQ ID NO:38)**

```

>gi|240173202|ref|ZP_04751860.1| FadD9 [Mycobacterium kansasii ATCC
12478]
MSTTRDERLERRIDTLIHDDAQFAAAKPDPAIAAALEKPGLSLPEIIQTLQGYADRPALGQRRAVEFVT
DTQTGRTSVRLITRFETITYRQLGDRVGALARALTHDSVHAGDRVCVLGFNSLDYTIDMALAKVGAVSV
PLQTSAAVTQLQPIVAETEPTMMAASVNQLSDAVDVLLSGHLPALKLVFDYHPEVDDQREALDTARERLA
DTAVVVQTLKDVLHDHGATLAAGSVAEPLAASGDNDISLALLIYTSGSTGAPKGAMYRQSNVGKMWRSSKN
WFGPTAACASITLNFMPPMSHIMGRGVLYGTGNGGTAYFAARSDLSTLLEDLRLVRPTELNFVPRIWETLYG
EYQRAVDQRSVDPGEPAAREAVEAQVMAEQRQDLGGRYIFAMTGSAPMSPELRNWEALLEIPLLDGYG
STEAGMVMFDGEIQRPPVIDYKLVDVPDLGYFSTDQPYPRGELLKTENMFPGYYKRPEVTASVFDADGY
YRTGDVVAEVAPDRLVYVDRRRNNVLKLAQGEFVTAKLEAVFGNSPLVRQIYVYGNSAHPYLLAVVPT
EASAGTDIAALKPLIADSLQTVAKEAGLQSYEVPRDFLIETPFTLENLGIRKLAWPKLRQHYGERL
EQLYTELAASQANELSELRSGAHAPVLETVSRAAGALLGAASTALSPDAHTDLGGDSLSAITFGNLLR
EIFDVDPVGVIVSPASDLAAIAAYIEGERQGSKRPTFAVIHGRDALEHVHASDLTLDKFIDASTLAAAPV
LPPPSAAVRTVLLTGTGFLGRYLALDWLERMDLVDGKVIALVRAKSDDARARLDKTFDSDGDELLTHY
RRLATDHLEVIAGDKGEANGLDQLTWQLADTVLIVDPAALVNHLPESELFGNPNALGTAELIRIALT
GKLKPYTYVSTIGVGDQIEPGKFTEDADIRHISATRKINDSYANGYGNISKWAGEVLLREAHDLGGLPVAV
FRCDMILADTTWAGQLNVPMFTRRMMISLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAEAIATLGA
RDGKGFQTYHVMNPYDDGIGMDRFDWLVDAGCAIHRIDDYGDWLRFETALRGLPEKQRHASLLPLLHN
YQKPAPPLRGSMAPTDRFRAAVQDAVKGPDKDIPHISPCIIAKYLSDLRLLGLL

```

**YP 978699.1****Nucleotide sequence (SEQ ID NO:39)**

```

>uniprot|A1KLT8|A1KLT8_MYCBP Probable fatty-acid-CoA ligase fadD9
ATGTCGATCAACGATCAGCGACTGACACGCCGCTCGAGGACCTATACGCCAGCGACGCC
CAGTCGCCGCCAGTCCCACAGCGATCACCCAGGGCTACGCCGATCGACCAGCCCCGGGTC
GCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTACGCCGATCGGCCGACTCGGC
CAGCGTGCCTCCGCTTCGTCACCGACCCGACAGCGGCCGACCATGGTCAGCTACTG
CCGCGGTTGAGACCATCACCTACCGCGAAGTGTGGCCCGCCGGCACATTGGCCACC
GCGTGAGCGTGAAGCCCGCATCCGGCGGGCGACCGGGTTTGCCTGGCTTCAAC
AGCGTCGACTACACAACCATCGACATCGCGTGTACCGGTTGGCGCCGTGCGTTCCA
CTGCAGACCAGTGCCTGGTCACCGGGTTGCCTGGCGCCGATCGTACCGAGACCGAGCCGACG
ATGATGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGTGCTGGCGGTACGCC
CCGGCCCGGCTGGCTATCGATTACCGACGGCAAGGGTACACCCACCGCGAGGCCGTC
GAAGCCGCCGAGCTCGGTTGGCCGGCTGGTACCGATCGACACACTTGCGAAGTGTAC
GAACCGCCGAGGGCGCTGCCGCCACACCCATTGCCGACAGCGCCGACGACGCCGCTGGCG
CTGCTGATTACACCTCGGGTAGTACCGCGCACCAAAGGGCGCATGTATCGCAGAGC
CAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGGTCAGGCCGAGCGGTTACCCCTCG

```

**FIG. 8T**

ATCACGCTGAACTTCATGCCATGAGCCACGTGGGGGCCGTCAGGTGCTACGGGACG  
CTTCCAACGGCGGTACCGCCTACTACGTCGCAAGAGCGACCTGTCGACGCTGTCAG  
GACCTCGCCCTGGTGCGGCCACAGAATTGTGCTCGTGC CGC GATCTGGGACATGGTG  
TTCGAGAGTTCCACAGCGAGGTGACCGCCGCTGGTGACGCCGATCGAGCG  
CTGGAAGCGCAGGTGAAGGCCGAGCTGCCGAGAAGCTGCTCGGCCGACGGTTGTCATG  
GCGCTGACGGGTTCCGCGCAGTCCTCGCTGAGATGACGGCGTGGTCGAGTCCCTGCTG  
GCCGACGTGCAATTGGTGGAGGGTTACGGCTCCACCGAGGCCGGATGGTCCTGAACGAC  
GGCATGGTGC GGCGCCCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGC  
TACTCGGCACCGATCAGCCCTACCCCCGGGCGAGCTGCTGGTCAGAACGCAAACCATG  
TCCCCCGGCTACTACCAGCGCCCGAGATGTCACCGCCGAGGTGTTGACCCCGACGGCTTC  
TACCGGACGGGGACATCATGGCAAAGTAGGCCGACCAAGTTCGCTACCTCGACCGC  
CGCAACAAACGTGCTAAAGCTCTCCAGGGCGAGTTCATGCCGTGCGAAGCTCGAGGC  
GTGTTGGCGACAGCCCCGCTGGTCGACAGATCTTCATCTACGGCAACAGTGCCGGGCC  
TACCCGCTGGCGGTGGTTGTCCTCGGGGACGCGCTTCTCGCCATGGCATCGAGAAT  
CTCAAGCCCGTGTAGCGAGTCCTGCAGGAGGTAGCGAGGGCGCCGCTGCAATCC  
TACGAGATTCCACCGCACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCTG  
CTCACCGGCATCCGCAAGCTGGCACGCCAGTTGAAGAAGTTCTATGGCGAACGTCTC  
GAGCGGCTCTATACCGAGCTGGCGATAGCAAATCAAACGAGCTGCGCAGCTGCGCAA  
AGCGGTCCCAGTGCGCCGGTGTCTCGACGCTGTGCCGTGCCGCGCTGCGTTGGC  
TCTACCGCTGGGATGTGCCGGGAGCGCAGCTCGCCGACCTGGGCTGGTACTCGCTC  
TCGGCGCTGCTGGTGGCCAACCTGCTGCAGAGATCTCGCGTGCACGTGCCGGTGGT  
GTCATTGTCAGCCCGAAGCGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGC  
ACCGGCGTCAGCGACCCAGCTCGCCTCGATAACCGTCGCTCCCGACGGAAGTGCAC  
GCCAGCGACCTCACGCTGGACAAGTTCATCGACGCTGCCACCTGCCGAGCCCCGAAC  
CTGCCGGCACCGAGGCCAAGTGCACCGTACTGCTGACCGGCCACCGGCTTTTG  
GGTCGCTACCTGGCGTGGAAATGGCTGACCGCATGGACCTGGTCAACGGCAAGCTGATC  
TGCCTGGTCCCGCCAGATCCGACGAGGAAGACAAGCCGGCTGGACGCGACGTTGAT  
AGCGGCACCGTATTGGTGGGCAACTACCGCAATTGGCGCCGGCCCTCGAGGTG  
CTGCCGGCACAGGGCAGGCCACCTGGGCTGGACCGGGTCAACCTGGCAGCGGCTA  
GCCGACACGGTGGACCTGATCGTGGACCCCGGGCCCTGGTCAACACAGTGTGCCGTAT  
AGCCAGCTGTCGGCCAAACCGGGGACCGCCGAGTTGCTCGGCTGGCGTGC  
GGCAAGCGCAAGCCATACATCTACACCTCGACGATGCCGTGGCGAGCAGATCCCG  
GAGCGTTCACCGAGGACGCCAGATCCGGGCACTCAGCCGACCCCGAGGATCGACGAC  
AGCTACGCAACGGCTACCGCAACAGCAAGTGGCGGGAGGTGCTGCGAAGCT  
CACGAGCAGTGC GGCTGCCGTGACGGTCTCCGCTGCGACATGATCCTGGCGACACC  
AGCTATACCGGTCAACCTGCCGGACATGTTCACCGGCTGATGCTGAGCCTGGCC  
GCTACCGGCATCGCACCCGGTCTGTTCTATGAGCTGGATGCGCACGGCAATCGCAACGC  
GCCCACTATGACGGCTTGCCTCGAATTGCGCAGAACGCCATTGCAACCTGGGACA  
CATAGCCGGACCGTTTGTCACCTACCGTGAACCCCTACGACGACGGCATCGGG  
CTGGACGAGTTGTCGACTGGCTCAACTCCCCAAGTAGCGGGTCCGGTTGCACGATCCAG  
CGGATCGCCGACTACGGCAGTGGCTGCAGCGGTTGAGACTTCGCTGCGTGCCTGCCG  
GATGCCAGCGCCACGCCCTCGCTGCTGCCCTGCTGCACAAGTACCGAGAGCCTGCAAAG  
CCGATATGCCGGTCAATCGGCCACCGACCAAGTCCCGCTGCCGCTCAAGAAGCGAAA  
ATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGATCATCGCAAGTACATCAGC  
AACCTGCGACTGCTGGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:40)**

>uniprot|A1KLT8|A1KLT8\_MYCBP Probable fatty-acid-CoA ligase fadD9

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIIDQPGVALPQLIRMVM  
EGYADRPALGQRALRFVTPDPSGRTMVELLPRFETITYRELWARAGTLAT  
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL  
RPIVTETEPTMIATSIDNLGDAVEVLGHAPARLVFDYHGKVDTREAV

**FIG. 8U**

EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG  
APKGAMYRESQVMSFWRKSSGWFEPSCGPSITLNFMPPMSHVGGRQVLYGT  
LSNGTAYYVAKSDLSTLFEDLALVRFTELCFVPRIWDMVFAFHSEVDR  
RLVDGADRAALEAQVKAELRENVLGGRFVMAUTGSAPISAEMTAWVESLL  
ADVHLVEGYGSTEAGMVLDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR  
GELLVKTQTMFPGYQRPDVTAEVFDPDGFYRTGDIAMAKVGPDQFVYLD  
RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIFIYGN SARAYPLAVVPSG  
DALSRHGIENLKVISESILQEVARAAGLQS YEIPRDFIETTPFTLENGL  
LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELROSGPDAPVLP  
LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDPVG  
VIVSPASDLRALADHIEARTGVRRPSFASTIHGRSATEVHASDLTLDKF  
DAATLAAAPNLAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI  
CLVRARSDEEAQARLDATFDGDPYLVRHYRELGAGRLEVLAGDKGEADL  
GLDRVTWQRЛАDTVDLIVDPAALVNHPYSQLFGPNAAGTAELLRLALT  
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAIISPTRRIDDSYANGYANSK  
WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA  
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH  
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP  
DRQRHASLLPLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPH  
TAAIAKYISNLRLLGLL

**ZP 05227804****Nucleotide sequence (SEQ ID NO:41)**

>gi|163719878:4480-7968 Mycobacterium intracellulare ATCC 13950  
NZ\_ABIN01000296, whole genome shotgun sequence

TTGGCCCGACCGACGAACAGTCCGCAACGCCAACCGATCTGTCGCTGCAGCAGGCGGCCAAC  
CGGGCCTGCGCCTGCCCAAGATTCTGGAGCTGTCGAGGGCTACGCCGATCGGCCGGTGGCTG  
GCGGCCAGGACACTGAGCACCGACCCCGGACCGGGCGTACCAACCACCGACTGCTCCCCCGCTTCGAC  
ACCATGACCTACCGCGAGCTGTGGCCGATGTGCGCGATCGCCGCCGCGTGGCGGACGACGCCGCGA  
ACCCGTGTCGCCCGGGGACTTCGTCGACGGTCGGTTCGCCAGCGCCAATATCTGACCCCTGACCT  
GGTCTGCGGCTACCTCGGGCTGGTTGCCGTTCCGCTGCAACACAATACGACCCCTCACGGCTGCGACCG  
ATCGTCGACGAGGTCGAGCCGTCGATACTGGCCCGCGCTGGTTATCTCGACCTCGCGTCGAGGCGG  
CGTGGCGAGCTCGTCTGCGCCGCTCGTGGCTTCGATTACGAGGTCGACGAGCAACGCGA  
GGCGCTGCAGCGCCAGGCACGCTGGCGCCGCCGGCGCGTGGACCATCGAGACCCCTGACGAA  
ATCATCGAACCGGGACGCCCTGCCGCCGAGCCATGTAACACGGTGATAACCGATCAGCGCTGGCGA  
TGATCATGTACACCTCGGGCAGCACCGGGTTACCAAGGGCGCCATGTACACCGAGCAGATGCTGGCAA  
GGTGTGGACCAACGAGCTGATGCCGACTTCGCGGACACACCCGTGTTCAACGTCACCTCATGCCGCTG  
AATCACCTCGTGGCCGGATACCGCTGTCGACCCGTTCCAGGCCGGCACCAGCTATTCTGTGCCGG  
AAAGCGACCTGTCACGTTGTCGACGACTGGAACCTGGTGCGCCAACCGAGATGGCCTGGTACCCCG  
GGTGGCGGAAATGCTTACCGCGCTACAGAGCGCCGTCGACCGACTCGTGGCTCGGGCGCCACGCC  
GGCTCCGCCAGGCCGCCGGCGCGAGCTCGTGAGCATGTCCTGGCGGGCGCATCGTACCGCCT  
TCTCGGGACGGCGCCGCTGGCGCGAGATGCGGGCCTCTGCGAAACCTGTTGGACGTCACGTTCT  
CGACGGCTACGGCTGACCGAGGTCGGCATGGTACCGAGCAAGGACGGCGCATGACCGTCCCCCGGTGCTC  
GACTACAAGCTACGACGTTCCGAACCTGGCTATTCCACACCGACAAGCCTTATCCGCGTGGCGAAT  
TGCTGGTGAAGTCGCTGACCGCGACGCCGGCTACTTCAAACGACCGACGTCACCGCAACCGTTCGA  
TCCCGACGGCTACTACCGGACCGCGATGTGATGGCGAGCTCGAGCCGGACCGGCTGGCCTACGTCGAC  
CGCCGCAACACGTTGAAGTGGCGAGGGCGAGTCGTCGCCGTCGCCGCTGGAGGCCGCTTC  
CCAGCGCGCCGCTGATCCGCCAGATCTCGTATACGGCAACAGCGAACGCCCTATCTGCTGGCGCTG  
CGTGGCGACGGCCGACGCCGCGAGCGATTACCGGAGATCCGAGGGCCTCAAGGCCGCGTGC  
TCCCTGCGCCAGTCGGCGCAACTCGCGAATCGCAGTCTACGAGGTGCCGTCGACTTCGTCGAGA

FIG. 8V

CCGAGCCGTTCAGCGAGGACAACGGCCTGCTCTGGCGTGGCAAGCTGCTGCCGAAGCTCAAGGA  
GCGCTACGCCAACCGGCTCGAACAGCTCACGCCAGCTGGCGAAAACCGCGTACCGAGTTGCGTGCG  
CTGCGCAGGGGGCGGACAAACACCCCGTCGCTTCACCCTCACCCGGCGCGAGGCCTACTGGGTG  
TGGCCGGCGGCCGCCGCCGCCGACGCACTGTTCATCGAACCTCGGCCGAGGGCGCTACTGGGTG  
CTTCTCCAACCTGCTGCGCACATCTTCGACGTGCGACGTGCCGGTGGGAATGATCACCGGCCGCGACC  
GACCTGGGCCAGCTCGCGAACATCGTGAATCCGAAACGCAAATCGGATCACGCCGCCACATTGCGGA  
CGGTGACGGACGCCGCCGCCGAGGTCCGCGCCGAGCTCACCTCGACAAGTTCATCGACGCCGAC  
GACCTGGCCGCCGACCGAACCTGCCGCCGCGACCGGCACACCCCACACGGTCTGCTGACCGGCCG  
AACGGCTACCTCGGCCGCTCCCTGGCCCTCGAATGGCTCGAGGCCCTGCCGAGACCGGCCGAAAGCTCG  
TCTCCATCGTCCGCGACGGACACCGCGGCCGTCAAACGGCTGGAGGCCGTTTCGACAGCGGGGA  
TCCGCAAGTTGCTGGAGCGGGTCCGGACGCTGGCCGCCGAGCACCTGGAAGTCATCGTGGCGACATCGGT  
GAGCCAATCTCGGCCTGGACCAAGCGACTTGGCAGCGCCTGGCCAGAGCGTGGATCTGATCGTCCACC  
CGGCCCGTTGGTCAACCAACGTGCTGCCGTACGACCAACTGTTGGTCCGAACGTGCTGGCACCGCCGA  
GTTGATCCGCTGGCGATCACGACGCCATCAAGCCGTCACCTATCTGTCGACCGTCCGCGTGGCGATG  
ACGGTCGATCCCCGGAGTTGCGGAAGACGGCAGATCCGCGGGTCAAGCGCGTACGCCGATCGACG  
ACAGCTACGCGAACGGTACGCGAACAGCAAGTGGCCGGTAGGGTGTGCTGCGTGGCGCACGACCT  
GTGCCGGTTGCCGGTCCGCGTCTCCGCTCCGACATGATCCTCCGCCACAGCCGGTATGCCGGCAGCTG  
AACGTGCCAGATGCCCTACCCGCTTGATGTTGCGCTGCTGACCAACCGGCATCGCCGCCGACCACGTTCT  
ACCGGACCGACGAACACGGAAACCGAGGCCGACTACGACGGCTGCCGACTTCGTCGCGATCGACG  
AGCGGTCAACCACGCTCGGCGAACAGATGGCGGCCAGGAATCCGGCGGGTACCGCTCTATGACGTGATG  
AACCCACACGACGACGGCGTCCCTGGACGTGTCGACTGGCTGATCGCCGCCGGACACGACATCC  
GGCGCATCGAGGACTATGACGAATGGCTGGCCGCTTCACCACGGCGCTCGCGCTTACCGGACAAGCA  
GCGCCAGCATTGGTGTGCGCTGCTGGACGCCCTACCGGAAACCGCGACGCCACTGCGGGGAGCGCG  
GCCCGGACCGTCTCCGCCACGCCGTGCGGACGCCAAATCGGTGCGGACGAGGACATCCGCACC  
TGTCGGCGGTGATCGACAAGTACGTCGCCGACCTACGCCGCTGGCTGGTGTAG

**Amino acid sequence (SEQ ID NO:42)**

>gi|254822803|ref|ZP\_05227804.1| putative long-chain fatty-acid--CoA  
ligase [Mycobacterium intracellulare ATCC 13950]

MAATDEQFRNAQPDSLQLQQAARQPGLRLPQILELFVEGYADRPRAVGRWARTLSTDPATGRTTLLPRFD  
TMTYRELWADVRAIAAAWRHDAANPVSPGDFVATVGFASAELYTLDLVCGYLGVLAVPLQHNTTPSRLRP  
IVDEVEPSILAAGVGYLDLAVEAASGSLLRRLVVFQPEVDEQREALQRAQATLAAAGAAVTIETLDE  
IIERGRALPPEPMYTGDTDQRLAMIMYTSGSTGLPKGAMYTEQMLAKVWTNELMPDFADTPVNVNFMPL  
NHLGGRIPLSTAFAQGTSYFVPESDLSTLFDLWNLVRPTEMGLVPRVAEMLYQRYQSAVDRLVASGADA  
GSAEARARAELREHVLCGRIVTACGTAPLAEMRAFVETCLDVHVLGDYGLTEVMVTKDGRMTRPPV  
DYKLIDVPELYFHDKPYPRGELLVKSLTATPGYFKRPDVTAANAFDPDGYRTGVDVMAELEPDRLAYVD  
RRNNVLKLAQGEFVAVARLEAVFASAPIRQIFVYGNNSERPYLLAVVPTADAAERTGDPEGLKAAVAE  
SLRQSAQLAELQSYYEVPDFVVETEPFSEDNGLLSGVGKLLRPKLKERYADRLEQLYAELAENRVTELRA  
LREGADKHPVVFTLTRAEEALLGVAGGPPAPDALFIELGGDSLSALTFSNLLRDIFDVDPVGMITGPAT  
DLGQLAEYVESERKSGSRRPTFATVHGRGAAEVRAAEELTLDKFIDATTLAAAPNLPRATGTPHTVLLGA  
NGYLGRFLALEWLERLAETGGKLVSIVRATDTAAVKRLEAVFDSDGPOLLERFRTLAAEHLEVIVGDIG  
EPNLGLDQATWQRLAQSVIDLIVHPAALVNHLVLPYDQLFGPNVVGTAELIRLAITTRIKPVTYLSTVAVAM  
TVDPGEFAEDGDIRAVSAVRPIDDSYANGYANSKWAEGVLLREAHDLGLPVAVFRSDMILAHSHRYAGQL  
NVPDAFTRLMFSSLTTGIAPTTFYRTDEHGNRAVAHYDGLPADFVAEAVTLGEQMAAEESGGYRSYDVM  
NPHDDGVSLDVFDWLIAAGHDIRRIEDEWLGRFTTALRALPDKQRQHSVLPLL DAYREPATPLRGAP  
APTDVFRHAVRTAKIGADEDIPHLSAALIDKYVADLRLGLV

**FIG. 8W****YP 001288539.1****Nucleotide sequence (SEQ ID NO:43)**

&gt;uniprot|A1QUM2|A1QUM2\_MYCTF Fatty-acid-CoA ligase fadD9

ATGTCGATCAACGATCAGCGACTGACACGCCGGTGCAGGGACCTATACGCCAGCGACGCC  
CAGTTCCGCCGCCAGTCCCACAGGCGATCACCCAGGGCATGCCAGCAGCCCCGGGTC  
GCGCTTCCACAGCTATCCGTATGGCATGGAGGGCTACGCCGATGCCGGACTCGGC  
CAGCGTGCCTCCGCTCGTACCGACCCGACAGCGGCCGACCATGGTCGAGCTACTG  
CCGCGGTTGAGACCATCACCTACCGCGAACCTGTGGGCCGCCGCOGGCACATTGGCCACC  
GCGTTGAGCGCTGAGCCCGCGATCCGGCCGGCGACCGGGTTTGCGTGTGGGCTTCAAC  
AGCGTCGACTACACAACCATCGACATCGCGCTATCGCGTGTACCGGGTGGCGCCGTCGGTTCCA  
CTGCAGACCAGTGCGCCGGTACCGGGTTGCGCCGATCGTACCGAGACCGAGCCGACG  
ATGATGCCACCAGCATCGACAATCTGGCGACGCCGTCGAAGTGCTGGCCGGTACGCC  
CCGGCCCGCTGGTCGTATTGATTACACGGCAAGGTTGACACCCACCGCGAGGCCGTC  
GAAGCCGCCGAGCTCGGTTGGCCGGCTCGGTGACCATCGACACACTGCCGAACGTGATC  
GAACGCGCAGGGCGCTGCCGCCACACCCATTGCCGACAGCGGCCGACGACGCCGCTGGCG  
CTGCTGATTTACACCTGGGTAGTACCGGGCACCCAAAGGCCCATGTATCGCGAGAGC  
CAGGGTGTGAGCTCTGGCGCAAGTCGAGTGGCTGGTCGAGCCGAGCGGTTACCCCTCG  
ATCACGCTGAACCTCATGCCGATGAGCCACGTCGGGGCCGTCAGGTGCTCTACGGGACG  
CTTTCACGGCGGTACCGCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTCCAG  
GACCTCGCCCTGGTGCAGGGCACAGAAATTGTGCTCGTGCAGCGCATCTGGGACATGGTG  
TTCGCAGAGTTCCACAGCGAGGTCGACCGCCGTTGGTGGACGGCGCGATCGAGCGCG  
CTGGAAGCGCAGGTGAAGGGCGAGCTGCGGGAGAACGTTGCTCGCGGACGGTTGTCATG  
GCCGCTGACCGGTTCCGCCGATCTCCGCTGAGATGACGGCGTGGTCGAGTCCCTGCTG  
GCCGACGTCATTTGGTGGAGGGTTACGGCTCCACCGAGGGCGGATGGTCTGAACGAC  
GGCATGGTGCAGGCCGCCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGC  
TACTTCGGCACCGATCACCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACATG  
TTCCCCGGCTACTACCAGCGCCGGATGTCACCGCCGAGGTGTCGACCCGACGGCTTC  
TACCGGACCGGGGACATCATGGCAAAGTAGGCCCCGACCAGTTCGTCGACCGC  
CGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTACGCCGTCGACAGCTCGAGGCG  
GTGTTGCGCAGAGCCGCTGGTCCGACAGATCTCATCTACGCCAACAGTGCCGGGCC  
TACCCGCTGGCGGTGGTTGTCGGCTCGGGGACCGCCTTCCTGCCATGGCATCGAGAAT  
CTCAAGCCCGTGTACCGCAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGCTGCAATCC  
TACGAGATTCCACCGCACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCTG  
CTCACCGCAGCCGATCCGCAAGCTGGCACGCCGAGTTGAGAACGTTCTATGGCAACGTC  
GAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCCGAGCTGCCGAA  
AGCGGTCCCAGTGCAGCCGGTGTCCGACCGCTGTGCGTGCAGCGGCTGCGTTGCTGGC  
TCTACCGCTGCGGATGTGCGGGACCGCAGCTCGCCGACTTCGCGACCTGGTGGACTCGCTC  
TCGGCGCTGCGTGGCAACCTGCTGCAAGAGATCTTCGGCTGACGTGCCGGTGGGT  
GTCATTGTCAGCCCGAAGCGACCTGCGGGCCCTGCCGACCATCGAACAGCAGCGCG  
ACCGGGCGTCAGCGACCCAGCTCGCCTCGATAACAGGTGCGTCCGCGACGGAAAGTCGAC  
GCCAGCGACCTCACCGCTGGACAAGTTACGCCGTCGACGCTGCCACCCCTGGCGACGCCGAAAC  
CTGCCGGCACCGAGCGCCCAAGTGCACCGTACTGCTGACCGGCCACCGGCTTTTG  
GGTCGCTACCTGGCGCTGGAAATGGCTGACCGCATGGACCTGGTCAACGGCAAGCTGATC  
TGCCTGGTCCCGCCAGATCCGACGAGGAAGCACAAGCCGGCTGGACCGACGTTGAT  
AGCGGCACCCGTTGGTGCAGCCGACTACCGCGAATTGGCGCCGGCCGCTCGAGGTG  
CTCGCCGGCAGAACAGGGCGAGGCCGACCTGGGCTGACCGGGCTACCTGGCAGCGGCTA  
GCCGACACGGTGGACCTGATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTAT  
AGCCAGCTGTTGGCCCAAACCGCGCCGGACCGCCGAGTTGCTCGGCTGGCGCTGACC  
GGCAAGCGCAAGCCATACATCACACCTCGACGATGCCGTTGGCGAGCAGATCCCGCCG  
GAGGCCTACCGAGGACGCCGACATCCGGGCCATCAGCCGACCCGAGGATCGACGAC

**FIG. 8X**

AGCTACGCCAACGGTACCGAACAGCAAGTGGCCGGGAGGTGCTGCTGCCGAAGCT  
 CACGAGCAGTGCAGCTGCCGGTACGGTCTTCGCTGCGACATGATCCTGGCGACACC  
 AGCTATACCGGTACGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTGAGCCTGCC  
 GCTACCAGCATCGCACCGGTTCTATGAGCTGGATGCGCACGGCAATCGGAAACGC  
 GCCCACTATGACGGCTGCCGGTCAATTGCTCGAGAAGCCATTGCACCCCTGGGACA  
 CATAGCCCAGGACCGTTTGTACACCTACCGTATGAAACCCCTACGACGGCATCGGG  
 CTGGACGAGTTCGTCGACTGGCTCAACTCCCCAAGTAGCGGGTCCGGTTGCACGATCCAG  
 CGGATCGCCGACTACGGCAGTGGCTGCAGCGGTTGAGACTTCGCTGCGTGCCTGCCG  
 GATGCCAGGCCACGCCCTCGTGCACACAACCTACCGAGAGCCTGCAAAG  
 CCGATATCGGGTCAATCGGCCAACCGACCAGTTCCGCTGCCGTCCAAGAAGCGAAA  
 ATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCAAGTACATCAGC  
 AACCTGCGACTGCTCGGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:44)**

>uniprot|A1QUM2|A1QUM2\_MYCTF Fatty-acid-CoA ligase fadD9

MSINDQRLLRRVEDILYASDAQFAAASPNEAITQAIIDQPGVALPQLIRMVM  
 EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT  
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAAPVTGL  
 RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVFDYHGKVDTREAV  
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG  
 APKGAMYRESQVMSFWRKSSGWFEPSEGPSITLNFMPPMSHVGGRQVLYGT  
 LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR  
 RLVDGADRAALEAQVKAELENVLGGRFVMAUTGSAPISAEMTAWVESLL  
 ADVHLVEGYGSTEAGMVNLNDGMVRPAVIDYKLVDVPELGYFGTDQPYPR  
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMKVGPDQFVYLD  
 RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSG  
 DALSRHGIENLKPVISESLSQEVARAAGLQSYEIPRDFIIETTPFTLENL  
 LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELQSGPDAPVLPT  
 LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDPVG  
 VIVSPASDLRALADHIEAARTGVRRPSFASIHRSATEVHASDLTLDKFI  
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI  
 CLVRARSDEEAQARLDATFDSDGPYLVRYRELGAGRLEVLAGDKGEADL  
 GLDRVTWQLRADTVLDLIVDPAALVNHLPLYSQFLGPNAAGTAELLRLALT  
 GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAIISPTRRIDDSYANGYANSK  
 WAGEVLLREAHEQCGLPVTFRCDMILADTSYTGQLNLPDMFTRLMLSIA  
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH  
 VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRADIYGEWLQRFETSLRALP  
 DRQRHASILLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPH  
 TAAIIAKYISNLRLLLG

**YP 953393.1****Nucleotide sequence (SEQ ID NO:45)**

>uniprot|A1T887|A1T887\_MYCVF Thioester reductase domain

ATGTCTACTGATACCCCGAAGACCGCCTTGCCCGCCATCGCGATCTGTACGCCACC  
 GACCCCGCAGTTGCCGCCAGCACCGACGCCATCTCCACGCCATCGATCAACCA  
 GGGACGCACCTGCCCGTCATCGTCAGACCGTCCTCGACGGATAACGCCAGCGGGCCCG  
 CTCGGACAGCGCGCGGTCCGTTCGTACCGACCCGCCACGGCAGAACCAACCGAG  
 CTGCTCCCCCGCTCGAGACCACACCTACGCCAGTTGTCACGCCGATCCACGCCGTC

FIG. 8Y

ACCGGCCGCGCTGACCGATGTGCATCCGGCGACCCGGTGGCGGTGCTGGGCTTCACCAGC  
ATCGACTACACCACCGTCGACATGGCTTGGCGATGCTGGCGCGGTGGCGGTACCACTT  
CAGACCAGCGCGCCGGCCACCAACCGTCCGGCCGATCGTCGCCAGACCGAACCGTGCG  
ATCGCATCGTCGGTCGATGCCCTACCGACGCCGTCGGCTGGCTCTCGACGCTCCCACC  
GTGACCCGCTTGTGGTCTTCGATCATCGGCCGGGTCGACGATCATCGCACGCCCTC  
ATCTCCCGAGCGACCGGTTGCGCGCCAACTCGCCGATCGAGGTGAGGACATTAC  
GACATCGTCGCTGGGGTTCGAAACTGCCGTACGCGCAATTCTCGGCCGACGGTGAC  
GCGCTGAGCCTGCTGATCTACACCTCCGGCAGCACCGGCGCACCAAGGGTGCGATGTAC  
CCCCAACACCTGGTCGCCAACTCATGGCGCGTTGGCCCGTCTCTGGGGCGACCTG  
GGGGTCTTCCCAGCAATACCGCTGAACCTCATGCCGATGAGCCACGTGATGGCCGCGGA  
CTGCTCTACGGCACGCTGGACGCCGGCACCGCTATTTCGCGGCCAGGAGCGATCTG  
TCGACGTTCTGGAGGATCTGCCCTGGTGCGCCGACGCAGCTGAGCTTCGTGCCGG  
ATCTGGGACACCATCCACGCCAGGTGTCGGCAGGAACCTCGAGGCCGGCGTCGGATGCG  
ACCGAGGTGATCGCCGATCTGCGCGGAGCCTGCTGGCGGCCGCTACGTACGGCGATG  
ACGGGCTCCCGCCGCTGTCACCGGAGATGCGGGCCTTCGTCGAGAACCTGCTCGACGTG  
CACCTGATCGACGGTACGGCTGACCGAGGCCGGCGGTGTCGACGCCGAGCTGAGCTTCC  
CAACGCCCGCCGGTATCGATTACAAGCTCGTCACGTCGCCGACCTCGGCTACTTCTCC  
ACCGACCGCCCCCATCCGCGCCGAGCTTCTCGTCAAATCCGAGACGCTGTTCCCGC  
TACTACAAACGCCCGACGTACCGCCGAGATGTTGACGAAGACGGCTACTACCGCACC  
GGCGACATCGTCGCCGAGACGGGGCCGACCAGCTGACCTATCTGGACCCGCCAACAAC  
GTCCTCAAACGTCGCAAGGGCAATTGTCACCGTCTCCGACTGGAGGCAGTGTCCG  
AACAGCCGCTGGTCCGCCAGATCTACGTCACGGCAACAGCCTGGCCGCTACCTGCTG  
GCTGTGGTCGTGCCACCGAAGCCGCGCTGGCCGGTGCACGCCAAAGCCGCTGTGCC  
GAGTCACCTCAGGATGTCGCAAGGCAGCCGGCTGCACTGCTACGAGATCCCCCGGAT  
TTCCCTCTCGAGACGACGCCGTTACGCTGGAGAACGGCTTGCTGACCGGCATCCGAAAA  
CTGGCCCGCCCGAGACTCAGAGAGCGTTACGGCAACAGCTGAGGCCCTACACCAG  
CTGTCGAAGAGCAGGCCGACGAGCTGCGGGAGCTGCGCCGCTCCGGCGAGAGCGTCCG  
GCGCTGGAAACCGTCGGACGCCGCCGGGCGCTGCTCGGCACCACCGCAGGCGAGCTG  
GAGCCGAGCGCCCACTTCACAGATCTGGCGGGGATTGCTGCTGCGCTGACCTTCGCC  
AACCTGTCGCGCACATCTTCGACGTCGACGTCGGCTCGGTGTCGTCAGCCGCC  
ACCGATCTGCAGGCCCTGCCACTACGTCGAGTCGCCGCCGCCACGGTCGGTGC  
CCCACCTCGAATCGGTGACCGGCATTGGGACGACCCGGGACCGAGGTGCATGCCGC  
GATCTGACGTTGGACGAATTGTCGACGCCGACCCCTGGCGACGCCGACGTTGCC  
GGACCGCGCGCCGAGGTCCGACCGCTCTGCTGACCGGGGCGACCGGCTTCCCTCGGCC  
TATCTCGCTCTCGAATGGCTTGAGCGGATGGCGCTGGTGCACGCCGAAAGCTGATCTGCC  
GTCCCGGCCAAAGACGATGCGCAGCGGGGTTGCGCTGGACAGCACGTTGACAGCGGG  
GACCCGGAGCTGCTGCCACTACCGACGGCTGGCAGCCGACCATCTCGAAGTGATGCC  
GGCGACAAGGCCGACGCCGATCTGGACTCGACGCCGGACGTCGACGCCGCTGCC  
ACCGTCGATCTGATCGTCGATCCGCCGCCCTGGTCAACCACGTCCTGCCGACGCC  
CTGTCGCCCGAACGATCGTCGACGATCGCGCAGCGCAGAATCGACGACAGCTAC  
GCCAACGGCTACGGCAACAGCAAGTGGGCCGGAGGGTGCCTGCGCAGGCCGACCGAT  
CTGTCGCCGCTGCCGTTGCGTGTCCGCTGCGACATGATCTGCCGACACCACCTAC  
GCCGGTCAGCTGAACCTGCCGACATGTTCACCGCCTGATCTCAGCTGGTGC  
GGCGTCGCCGCCGAGTCGTTCTACCAACCTGCCACCGACGGCACCCGGCAACGGGCC  
TATGACGGGCTACCGGTGGAGTTATCGCGAAGCGATCTCGACCCCTCGGCTCTGACGTC  
GCCTCCGGGTTCCGGACGATCACGTGATGAATCCCCACGACGACGGGATAGGCTCGAC  
GAGTACGTTGACTGGCTATCGACGCGAGGCCATCCGATTGACGCGTCGGCGACTACCG  
ACGTGGCTGCAACGGTTACGGTGGCGATCACCGCCTGCCGGAAACGGCAGCGGCC  
TCAGTGGCTGCAACGGTTACGGTGGCGATCACCGCCTGCCGGAAACGGCAGCGGCC  
GCACCGACAGACAGATTCCGTAAGCGGTCCAGGACGCCAAATCGGTCCGGACAAAGAC  
ATTCCACACGTGACACCGCAGATCGTATCAAGTACGTACCGACCTGCAGCGACTCGGA  
CTGCTCTAA

**FIG. 8Z****Amino acid sequence (SEQ ID NO:46)**

>uniprot|A1T887|A1T887\_MYCVP Thioester reductase domain

MSTDTREDRLARRIADLYATDPQFAAAAPDDAISHAIDQPGTHLPVIVQT  
VLDGYAERPALGQRAVRFVTDPATGRTTTELLPRFETITYAELSRRIHAV  
TAALTDVHPGDRVAVLGFTSIDYTTVDMALAMLGAVAVPLQTSAPATTVR  
PIVAETEPVVIASSVDALTDAGLADAPTVRTLVVFHRAGVDDHRDAL  
ISASDRLRAANSPIEVETITDIVARGSKLPVRAQFSADGDALSLLIYTSG  
STGAPKGAMYPQHLVANSWRRLARSFWGDLGVFPATLNFMPPMSHVMGRG  
LLYGTLDAGGTAYFAARSDLSTFLEDLALVRPTQLSFVPRIWDTIHAEV  
QELEERRPSDATEVIADLRRSLLGGRYVTAMTGSAPLSPEMRAFVENLLDV  
HLIDGYGSTEAGAVFVDGRVQRPPVIDYKLVDVADLGYFSTDPRPHPRGEL  
LVKSETLFPGYYKRPDVTAEMFDDEGYRTGDIVAETGADQLTYLDRRNN  
VKLSQGEFTVSRLEAVFGNSPLVRQIYVYGNSSARPYLLAVVVPTEAAL  
AGADAKAAVAESLQDVAKATGLQSYEIPRDFLLETPFTLENGLLTGIRK  
LARPRLRERYGEQLEALYTMSEEQADELRELRRSGGERPALETVGRAAG  
ALLGTTAGELEPSAHTDLGGDSLSALTTFANLLRDI FDVDVPVGIVSPA  
TDLQALADYVESARRHGSVRPTFESVHGHSRPGTEVHARDLTLDEFVDA  
ATLAHAPTLPGPRAEVRTVLLTGATGFLGRYLALEWLERMALVGGKLICL  
VRAKDDAAARVRLDSTFDSDGDELLRHYRRLAADHLEVIAGDKADADLGL  
DARTWQLRADTVDLJVDPAALVNHVLPYRQLFAPNVLGTAELLRIALTTR  
MKPFVYVSTIGVGAGIEPARFTEDADIRQISATRRI DDSYANGYGN SKWA  
GEVLLREAHLCGLPVSVRCMDMILADTTYAGQLNL PDMFTRLIFSLVAT  
GVAPESFYHLATDGTRQRAHYDGLPVEFIAEAISTLGSVASGFRTYHVM  
NPHDDGIGLDEYVDWLIDAGHPIRRVGDYPTWLQRTVAITALPERQRQA  
SLLPLLHNQH PETPIRGSIAPTDRFREAVQDAKIGPDKDIPHVT PQIVI  
KYVTDLQRLGLL

**YP 938306.1****Nucleotide sequence (SEQ ID NO:47)**

>uniprot|A1UFA8|A1UFA8\_MYCSK Thioester reductase domain

ATGTCCACCGAGACCCGTGAAGCGCGCTGCAGCAGCGCATCGCGCATCTGTT CACCACC  
GACCCCGCAGTCGCCGCCGCCGCCGCCGACCCCCGGATCAGCGATGCCGTGACCGCGAT  
GACACGCGGCTGACCGCCATCGTCTCGCGGTATGTCGGGGTACGCCGACCGGCCGGCA  
CTCGGGCAGCGGCCCGAATT CGTACCCGACCCCGAGACCGGCCGACACGATGGAG  
CTGCTCCCCCGCTCGACACCATCACCTACCGGGAGCTGCTGACCGCGTCCGGCGCTC  
ACCAACCGTGGCACGCCGACGGGGTGCGCCCGGCGACCGTGTGCAATCCTCGGATTC  
ACCGGCATCGACTACACCGTGGTCGACCTCGCCCTGATCCAGCTCGCGCGTGCCTG  
CCGCTGAGACCAGCGCCGCCGTGGAGGC GTGGCGCCGATCGTGGCGAGACCGAACCG  
ATGCTCATCGCCACCGCGTCGATCATGTCGACGCCGCCGCGGA ACTCGCACTCACCGGC  
CACCGTCCGTCCC GG GTGGTGGTCTCGACCCACCGCGAGCAGGTGACGACGAACCGAC  
GCGGTGCGGGCGCGACGCCAGGGCTGGAGACGCCGTCCCCGTGAGAGACACTCGCCGAG  
GTGTTGCGGC GGGGCCCATCTGCCCGCCGTGGCGCCGACGTGTTGACGAGGCCGAT  
CCACTGCGGCTGCTGATCTACACTTCCGGCAGCGCCGGCGCCCCAAGGGCGCGATGTAT  
CCCGAGAGCAAGGTGCGCCGGCATGTGGCGCGTCGGCCAAGGGCGCTGGAACAACGAT  
CAGACAGCGATTCCGTCGATCACCCCTGAACCTCCTGCCGATGAGCCACGT CATGGGT CGC  
GGCCTGCTGTGCGGCACGCTCAGCACCGGTGGCACCGCGTATT CGCCGCCCGCAGCGAT  
CTGTCGACGCTGCTCGAGGACCTGCGCCTGGTACGGCCCACCCA ACTCAGCTCGTCCG

**FIG. 8AA**

CGGATCTGGACATGCTCTTCAGGAGTCGTGGCGAGGTGACCGGGGGTGAACGAC  
GGTGC GGACGCCAACCGCGAGGCCGACGTGCTGGCGTACAGCGTCACGAGCTGCTC  
GGTGGCGGTTCGTCACCGCGATGACC GGTTCGGCGCCATCTCCCTCGAGATGAAGACA  
TGGGTGGAGACCTGCTCGACATGCACCTGGTCGAGGGTTACGGCTCGACGGAGGCC  
GCGGTGTTCGTCGACGGCCACATCCAGCGCCACCGGTGCTGACTACAAACTCGTCGAC  
GTCCCCGACCTCGGCTACTTCAGCACCGACCGGCCACCGCGCAGCGGTGAGCTGCTGGTC  
CGCTCCACCGCAGCTATTCCCCGGCTACTACAAACGTCCCGACGTACCGCCGAGGTGTT  
GACGACGACGGCTTCTACCGCACGGCGACATCGTCGCCAGGTGCGCCCCGATCAGGTG  
CA GTACCTCGACCGCCGAACAAACGTGCTCAA ACTCGCCCAGGGTGAGTTCGTCACCAC  
TCCAAACTCGAGGCGGTCTCGCCGGCAGCGCCCTGGTCCGCAAGATCTACGTGTACGGC  
AACAGTGC CGCCTCACCTGCTGGCGTGTGCGACCGACGATGCGGTGGCC  
CACGACCCGGCATCGCTCAAGACCGCGATCAGCGCTCGTGCAGCAGGCCGCGAAGACC  
GCCGGTCTGCAGT CCTACGAGCTGCCCGTGACTTCCCTCGTCGAGACTCAACGTTACG  
CTGGAGAACGGACTACTGACCGGCATCCGCAAGCTGGCGGCCGAAACTCAAGGCGCGC  
TACGGCGATCGGCTCGAGGC GTCTACGTGCAACTGGTCAAGGACAGGCAGGCGAATTG  
CGCACCCCTCGCCGGGACGGCGCGAAGCGTCCGGTGGCGAGACGGTGGCCCGCG  
GCCCGCCTGCTCGGCCGCCGCGCGACGTGCGCCCCGACCGCACTTCACCGACCTC  
GGCGGAGACTCGTTGTCGGCGTTGACCTT CGGCAATCTGCTGCAAGGAGATCTCGGCGTC  
GACGTCCCGTGGGGT GATCGTCAGCCCGGGCGACCTGGCGTGTGATCGCGCGTAC  
ATCGAGGCCGAACAGGCCTCGACCGGTAAGCGGCCGACCTACCGCTCGGTGCACGGCGC  
GACGCCGAACAGGTACACGCCCGCGACCTCACCCCTGGACAAGTTCATCGACGCCGAAACA  
CTCTCCGCTGCAACAGAACTGCCCGCCCGAGCGGTGAGGTGCGCACCGTGTGCTGACC  
GGGGCCACCGGATT CCTCGGCCGCTAC TGGCGCTGGACTGGCTCGAACGGATGGCCCTG  
GTCGACGCCGAAGGTATCTGCCCTGGCGCGAAGGACGATGCCGCCGCAAACGC  
CTCGACGACACCTCGACAGCGCGATCCGAAGCTGCTGGCGCACTACCGCAAGCTGCC  
GCCGACCACCTCGAAGTGTGCGCCGGCGACAAGGGTGAGGC GGATCTGGCGTGC  
CCTGTCGGCAGCGCCTGGCCGACACCGTCGACCTCATCGTCGACCCGGCGCCCTGGTC  
AACACGTACTGCCGTACGCCAGCTGTTGGGCCAACCGCGTGGGCAC TGCGAGCTG  
ATCCGGCTTGC GCTACCACCCGATCAAACCGTTACCTACGTGTCGACGATCGGCGTC  
GGCGCCGGTATCGAACCGGGCGTTT CACCGAGGACGACGACATCCGGGTGATCAGCCG  
ACCGGGCCGTCGACACCGGGTACGCCAACGGTACGCCAACAGCAAGCAAGTGGCCGGTGAG  
GTGTTGTTGCGCGAGGCGCACGATCTGTGCGGGCTCCCGTGGCGTGTCCGGTGC  
ATGATCCTGGCCGACACCACCTACGCCGGCCAGCTCAACCTGCGGACATGTTACCCGG  
ATGATGCTGAGCCTGGT GACCACGGGTATCGCGCGAAATCGTCCACCCACTCGAC  
AAGGGCCACCGGCAAGGCCCACTACGACGGGCTGCCGGTCGAGTTGCTGCCGAATCG  
ATCTCCGCGCTGGGAGCGCAGGCCAGGGATCGGCTCGACGAATTGCTGACTGGCTCG  
GTGATGAACCCCACGACGACGGGATCGGCTCGACGAATTGCTGACTGGCTCG  
GCCGGGTATCGCATCGACCGCATCGACTACTACGCCGCTGGCTGCAGCGGTTG  
GCCCTGCCGGCGCTGCCGAGCGCACTGCCAGTACTCACTGCTCCCGTGTGC  
TACCAAGCGGCGCTGCCGACCGGATCAACGGGGGATGCCCGGACCGACCGGTTCC  
GCGGTGCAAGGAGGCAAAGCTGGCCCGGACAAGGACATTCCCACGTCACTCC  
ATCGTCAAGTACGCCACCGACCTGGAGCTGCTGGCGTGT  
G

**Amino acid sequence (SEQ ID NO:48)**

>uniprot|A1UFA8|A1UFA8\_MYCSK Thioester reductase domain

MSTETREARLQQRIAHLFDTDPQFAAARPDPRI SDAVDRDDTRLTAIVSA  
VMSGYADRPALGQRAAEFVTDPQTGRTTMELLPRFDITIYRELLDRV  
TNAWHADGVPGDRVAILGFTGIDYT VVDLALIQLGAVAVPLQTS  
LRPVAE TEPM LIATGVDHVAAAELALTGH RPSRVVFDHREQV  
DERRD AVRAATARLGDAVPVETLAEVILRRGAHLPAVAP  
HFDEADPLRLLIYTSG  
SAGAPKGAMYPESKVAGMWRASAKA AWWNNDQTAIP  
SITLNFLPM SHVMGR  
GLLCGTLSTGGTAYFAARSDLSTLLEDLRLVRPTQLSFV  
PRIWDMLFQE F

**FIG. 8BB**

VGEVDRRVNDGADRPTAEADVLAVQRHELLGGRFVTAMTGSAPISLEMKT  
WVETLLDMHLVEGYGSTEAGAVFVDGHIQRPPVLDYKLVDVLDI.GYFSTD  
RPHPRGELLVRSTQLFPGYKRPDVTAEVFDDDFYRTGDIVAEVGPDQV  
QYLDRRNNVLKLAQGEFVTISKLEAVFAGSALVRQIYVYGNARSYLLAV  
VVPTDAVARHDPAISLKAISLQQAAKTAGLQSYELPRDFLVEQFQFT  
LENLITGIRKLARPKLKARYGDRLEALYVELVEQAGELRTLRRDGAKR  
PVAETVGRAAAALLGAAAADVRPDAHFTDLGGDSLSALTFGNLLQEIFGV  
DVPVGIVSPAADLASIAAYIEAEQASTGKRPTYASVHGRDAEQVHARDL  
TLDKFIDAETLSAATELPGPSGEVRTVLLTGTGFLGRYLALDWLERMAL  
VDGKVICLVRACKDAAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLAGD  
KGEADLGLPHPWQRLADTVLIVDPAALVNHVLPYSQLFGPNALGTAEL  
IRLALTTRIKPFTYVSTIGVAGIEPGRFTEDDIRVISPTRAVDTGYAN  
GYGNSKWAGEVLLREAHDLCLGLPVAVFRCDMILADTTYAGQLNLPMFTR  
MMMSIVTTGIAPKSFHPLDAKGRQSAHYDGLPVEFVAESISALGAQAVD  
EAGTGFATYHVMNPHDHGIGLDEFVDWLVEAGYRIDRIDYYAAWLQRFET  
ALRALPERTRQYSLLPLLHNYQRPAHPINGAMAPTDRFRAAVQEAKLGPD  
KDIPHVTPAVIVKYATDLELLGLI

**ZP 05217435**

**Nucleotide sequence (SEQ ID NO:49)**

>gi|222089526:2534-6055 Mycobacterium avium subsp. avium ATCC 25291  
NZ\_ACFI01000138, whole genome shotgun sequence

ATGTCGACTGCCACCATGACGAACGACTCGACCCTCGCACGAACTCATGCCACCGACCCGCAAT  
TCGCCGCCACCCACCCGACCCGGCGATCACCCTCGAACAGCCGGGCTGCCGCTGCCAGAT  
CATCCGCACCGTGCTCGACGGCTACGCCGACCGGGCGCTGGGACAGCGCGTGGAGTCGTCACG  
GACGCCAAGACCGGGCGCACGTGGCGCAGCTGCTCCCCCGCTTCGAGACCATCACGTACGGCGAAGTGG  
CGCAGCGTGTTCGGCGCTGGCGCCCTGTCCGACGACGGGTGACCCCGGGGACCGGGTGTGCGT  
GCTGGGCTTCACAGCGTCAACTACGCCACCATCGACATGGCGCTGGCGCCATGGCGCCGTCTCGGTG  
CCGCTGCAGACCAGCGCGGAATCAGCTCGCTGAGCCGATCGTGGCCGAGACCGAGCCCACCTGATCG  
CGTCCAGCGTGAACCAAGCTGTCCGACCGGGTGCAGCTGATCACCGCGCCGAGCAGGCGCCACCCGGCT  
GGTGGTGTGACTACCAACCCGAGTCGACGACCAAGCGGAGGGCGTCCAGGACGCCGGCGGGCTG  
TCCAGCACCGGGTGGCGTCCAGACGCTGGCGAGCTGCTGGAGCGCGGAAGGACCTGCCCGCTCG  
GGGAGCCGCCCGACGAGGACTCGCTGGCGTCTGATCTACACCTCCGGGTCACCGGGCCCCCAA  
GGCGCGATGTACCCGAGAGCAACGTGGCAAGATGTGGCGCCGGCAGCAAGAACTGGTTCGGCGAG  
AGCGCGCGTCACTTACCCCTGAACTTACCGTGGCGTGGCGAGCTGAGCCACGTGATGGCGAAGCATCTACGGCA  
CGCTGGCAACGGCGGACCCGCTACTTCGCCGCCGAGCGACCTGTCGACGGGCTCCACCGTCTCGAGGACCTCGA  
GCTGGTGGGCCACCGAGCTCAACTTCGTCCCGGGATCTGGAGACGCTGTACGGCGAATTCCAGCGT  
CAGGTCGAGCGGGCTCTCGAGTCCGGGACGCCGGCAACGTGCGCCGTGAGGCCGAGGTGCTGG  
CCGAGCAGGCCAGTACCTGCTGGCGGGCGTTACCTCGCGATGACGGGCTGGCGCCATCTCGCC  
GGAGCTGCGCAACTGGGTGAGTCGCTCGAAATGCACCTGATGGACGGCTACGGCTCCACCGAGGCC  
GGAATGGTGTGTTGACGGGAGATTCAAGGCCGCCGGTGTGACTACAAGCTGGTCAACGTGCCGG  
ACCTGGGCTACTCAGCACCGACCGGGCGCATCCGCGGGGAGCTGCTGTCGCCACCGAGAACATGTT  
CCCGGGCTACTACAAGCGGGCGAAACCCACCGCGGGCGTCTCGACGAGGACGGCTACTACCGCACCGGC  
GACGTGTTGCCGAGATCGCCCGGGACCGGCTGGTCTACGTCGACCGCCGCAACACGTGCTCAAGCTGG  
CGCAGGGCGAATTGTCACCGCTGGCCAAGCTGGAGGCGGTGTTGGCAACAGCCCGCTGATCCGCCAGAT  
CTACGTCTACGGCAACAGCGCCCAGCCCTACCTGCTGGCGGTGTCGGTGCACCGGAGGCC  
TCGGGTGACCCCGAGACGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTGCCAAGGAGGCC  
TGCAGTCTACGAGGTGCCGCGACTTCATCATCGAGACCACCCGTCAGCTGGAAAACGGTCTGCT  
GACCGGGATCCGGAAGCTGGCGTGGCCGAAACTGAAGCAGCACTACGGGAACGGCTGGAGCAGATGTAC  
GCCGACCTGGCCGCCGGACAGGCCAACGAGCTGGCGAGCTGCCGCAACGGTGCCAGGCCGGTGC

**FIG. 8CC**

TGCAGACCGTGAGCCGCCGCGGGGCCATGCTGGGTCGGCCCTCCGACCTGTCCCCGACGCCA  
 CTTCACCGATCTGGCGGAGACTCGTTGTCGGCGTTGACATTGGCAACCTGCTGCGCGAGATCTTCGAC  
 GTCGACGTGCCGGTGGCGTCATCGTCAGCCC GGCAACGACCTGGCGCCATCGCGAGCTACATCGAGG  
 CCGAGCGCAGGGCAGCAAGCGCCCGACGTTCGCCTCGGTGACGGCCGGAGCGGACCGTGGTGC  
 CGCCGACCTGACGCTGGACAAGTTCTCGACGCCGAGACGCTGGCGCCGGCGAACCTGCCAAGCCG  
 GCCACCGAGGTGCGCACCGTGCTGACCGGCCACCGGCTTCTGGCCGCTACCTGCCCTGGAAT  
 GGCTGGAGCGGATGGACATGGAGGGCAAGGTCACTGCCCTGGTCCGGCCCTCCGACGAGGAGGC  
 ACGCCGCCGGCTGGACAAGACCTCGACAGCGCCGACCCGAAGCTGCTCGCAGCTACCAGCAGCTGGCC  
 GCCGATCACCTGGAGGTCACTGCCCGACAAGGGCGAGGCCAATCTGGGCTGGCCAAGACGTTGGC  
 AACGACTGGCCGACACGGTCGACGTGATCGTCGACCCGCCGCTGGTCAACCACGTGTTGCCGTACAG  
 CGAGCTGTTGGGCCAACGCCCTGGCACCGCGAGCTGATCCGGCTGGCGTGACGCTAACGAGAAG  
 CCGTACACCTACGTGTCCACCATGGCGTGGCGACCAAGATCGAGCCGGCAAGTCTGTCGAGAACGCCG  
 ACATCCGGCAGATGAGCGCCACCCGGCGATCAACGACAGCTACGCCAACGGCTACGGCAACAGCAAGT  
 GGCCGGCGAGGTGCTGCTGCGGAGGGCGACGACCTGTGCGGCTGCCGTGCGGTGTTCCGCTGCGAC  
 ATGATCCTGGCCGACACCACTGATGCCGGCAGCTCAACCTGCCGGACATGTTCACCCGGCTGATGCTGA  
 GCCTGGTGGCCACCGGGATCGGCCCGCTCGTCTACGAGCTGACGCCGACGGCAACCGGAGCGGGC  
 GCACTACGACGGCTGCCGGTCGAGTTCATGCCCGGCGATCTCGACGCTGGGTCGAGATCACCGAC  
 AGCGACACCGGCTCCAGACCTACCACTACGATGACGGCATCGGTCTGGACGAGTACG  
 TCGATTGGCTGGTGGACGCCGGTATTGATCGAGCGGATTGCCGACTACTCCGAATGGCTGCCGGTT  
 CGAGACCTCGTGGCCGGTGGACGCCGGTATTGATCGAGCGGATTGCCGACTACTCCGAATGGCTGCCGGTT  
 CGCACGCCGGAGAAAGCCGATCAACGGGTCGATAGCTCCACCGACGTGTTCCGGGAGCGGTGCGAGGAGG  
 CGAAAATCGGCCCGACAAAGACATTCCGACGTGTCGCCGCCGGTATCGTAAGTACATCACCGACCT  
 GCAGCTGCTGGCGTGTCTGA

**Amino acid sequence (SEQ ID NO:50)**

>gi|254775919|ref|ZP\_05217435.1| FadD9 [Mycobacterium avium subsp. avium ATCC 25291]

MSTATHDERRVHELIATDPQFAAAQPDPAITAALEQPGLRLPQIIRTVLDGYADRPALGQRVVEFVT  
 DAKTGRRTSAQLLPRFETITYGEVAQRVSALGRALSDDAVHPGDRVCVLGFNSVDYATIDMALGAI  
 GAVSVPLQTSAIASSLQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTRIVVFDYHPQVDDQREAVQ  
 DAAARLSSTGVAVQTLAELLERGKDLPAVGEPPADEDSLALLIYTSGTGAPKGAMYPQS  
 NVGKMWRRGSKNWFGE  
 SAASITLNFMPPMSHVMGRSILYGTGNGGTAYFAARSDLSTILLEDLELVRPELNFV  
 PRIWETLYGEFQR  
 QVERRLSES  
 GDA  
 GERRAVEAEVLA  
 EQRQYLLGGRFTFAM  
 TGSAPISPELRN  
 WVESLLEM  
 HLM  
 DGYGSTE  
 GMVLF  
 DGEIQR  
 P  
 PVVIDYKL  
 DVLP  
 DLGYF  
 STDRPH  
 PRG  
 ELLR  
 TENMF  
 PG  
 YYKRA  
 ET  
 TAGVF  
 DE  
 DGY  
 YRTG  
 DVFAE  
 IAP  
 DRL  
 VV  
 DRR  
 NN  
 VL  
 KLA  
 Q  
 GE  
 F  
 V  
 TL  
 AK  
 LE  
 AV  
 F  
 G  
 N  
 S  
 P  
 L  
 I  
 R  
 Q  
 I  
 Y  
 V  
 Y  
 G  
 N  
 S  
 A  
 Q  
 P  
 Y  
 L  
 L  
 V  
 V  
 P  
 T  
 E  
 E  
 A  
 L  
 A  
 S  
 D  
 L  
 Q  
 Q  
 V  
 A  
 K  
 E  
 A  
 G  
 L  
 Q  
 S  
 Y  
 E  
 V  
 P  
 R  
 D  
 F  
 I  
 I  
 E  
 T  
 T  
 P  
 F  
 S  
 L  
 E  
 N  
 G  
 L  
 L  
 T  
 G  
 I  
 R  
 K  
 L  
 A  
 W  
 P  
 K  
 L  
 K  
 Q  
 H  
 Y  
 G  
 E  
 R  
 L  
 E  
 Q  
 M  
 Y  
 A  
 D  
 L  
 A  
 A  
 Q  
 N  
 A  
 E  
 L  
 R  
 R  
 N  
 G  
 A  
 Q  
 P  
 V  
 L  
 Q  
 T  
 V  
 S  
 R  
 A  
 A  
 G  
 A  
 M  
 L  
 G  
 S  
 A  
 A  
 D  
 L  
 S  
 P  
 D  
 A  
 H  
 F  
 T  
 D  
 L  
 G  
 G  
 D  
 S  
 L  
 A  
 L  
 T  
 F  
 G  
 N  
 L  
 L  
 R  
 E  
 I  
 F  
 D  
 V  
 D  
 V  
 P  
 G  
 V  
 I  
 V  
 S  
 P  
 A  
 N  
 D  
 L  
 A  
 A  
 I  
 S  
 Y  
 I  
 E  
 A  
 E  
 R  
 Q  
 G  
 S  
 K  
 R  
 P  
 T  
 F  
 A  
 S  
 V  
 H  
 G  
 R  
 D  
 A  
 T  
 V  
 V  
 R  
 A  
 A  
 D  
 L  
 T  
 L  
 D  
 K  
 F  
 L  
 D  
 A  
 E  
 T  
 L  
 A  
 A  
 P  
 N  
 L  
 P  
 K  
 P  
 A  
 T  
 E  
 V  
 R  
 T  
 V  
 L  
 L  
 T  
 G  
 A  
 T  
 G  
 F  
 L  
 G  
 R  
 Y  
 L  
 A  
 E  
 W  
 L  
 E  
 R  
 M  
 D  
 V  
 D  
 G  
 K  
 V  
 I  
 A  
 L  
 V  
 R  
 A  
 R  
 S  
 D  
 E  
 E  
 A  
 R  
 A  
 R  
 L  
 D  
 K  
 T  
 F  
 D  
 S  
 G  
 D  
 P  
 K  
 L  
 L  
 A  
 H  
 Y  
 Q  
 Q  
 L  
 A  
 D  
 H  
 L  
 E  
 V  
 I  
 A  
 G  
 D  
 K  
 G  
 E  
 A  
 N  
 L  
 G  
 L  
 Q  
 D  
 V  
 W  
 Q  
 R  
 L  
 A  
 D  
 T  
 V  
 D  
 V  
 I  
 V  
 D  
 P  
 A  
 A  
 L  
 V  
 N  
 H  
 V  
 L  
 P  
 Y  
 S  
 E  
 L  
 F  
 G  
 P  
 N  
 A  
 L  
 G  
 T  
 A  
 E  
 L  
 I  
 R  
 L  
 A  
 L  
 T  
 S  
 K  
 Q  
 K  
 P  
 Y  
 T  
 Y  
 V  
 S  
 T  
 I  
 G  
 V  
 G  
 D  
 Q  
 I  
 E  
 P  
 G  
 K  
 F  
 V  
 E  
 N  
 A  
 D  
 I  
 R  
 Q  
 M  
 S  
 A  
 T  
 R  
 A  
 I  
 N  
 D  
 S  
 Y  
 A  
 N  
 G  
 Y  
 G  
 N  
 S  
 K  
 W  
 A  
 G  
 E  
 V  
 L  
 L  
 R  
 E  
 A  
 H  
 D  
 L  
 C  
 G  
 L  
 P  
 V  
 A  
 V  
 F  
 R  
 C  
 D  
 M  
 I  
 L  
 A  
 D  
 T  
 T  
 Y  
 A  
 G  
 Q  
 L  
 N  
 L  
 P  
 D  
 M  
 F  
 T  
 R  
 L  
 M  
 L  
 S  
 L  
 V  
 A  
 T  
 G  
 I  
 A  
 P  
 G  
 S  
 F  
 Y  
 E  
 L  
 D  
 A  
 D  
 G  
 N  
 R  
 Q  
 R  
 A  
 H  
 Y  
 D  
 G  
 L  
 P  
 V  
 E  
 F  
 I  
 A  
 A  
 I  
 S  
 T  
 L  
 G  
 S  
 Q  
 I  
 T  
 D  
 S  
 T  
 G  
 F  
 Q  
 T  
 Y  
 H  
 V  
 M  
 N  
 P  
 Y  
 D  
 D  
 G  
 I  
 G  
 L  
 D  
 E  
 V  
 V  
 D  
 W  
 L  
 V  
 D  
 A  
 G  
 Y  
 S  
 I  
 E  
 R  
 A  
 D  
 Y  
 S  
 E  
 W  
 L  
 R  
 R  
 F  
 E  
 T  
 S  
 L  
 R  
 A  
 L  
 P  
 D  
 R  
 Q  
 R  
 Q  
 Y  
 S  
 L  
 L  
 P  
 L  
 H  
 N  
 Y  
 R  
 T  
 P  
 E  
 K  
 P  
 I  
 N  
 G  
 S  
 I  
 A  
 P  
 T  
 D  
 V  
 F  
 R  
 A  
 V  
 Q  
 E  
 A  
 K  
 I  
 G  
 P  
 D  
 K  
 D  
 I  
 P  
 H  
 V  
 S  
 P  
 P  
 V  
 I  
 V  
 K  
 Y  
 I  
 T  
 D  
 L  
 Q  
 L  
 L  
 G  
 L  
 L

**FIG. 8DD**

YP 001070587.1

### Nucleotide sequence (SEQ ID NO:51)

>uniprot|A3PYW9|A3PYW9 MYCSJ Thioester reductase domain

ATGTCACCAGAGACCGCTGAAGCGCGCTGCAGCAGCGCATCCGCATCTGTCGCCACC  
GACCGCAGTCGCCGCCCGACCCCCGGATCAGCGATGCCGACCGCGCAT  
GACGCGGGCTGACCGCCATCGTCTCCGCCGTATGTCGGGTACGCCGACCGGCCGCA  
CTCGGGCAGCGCCGCGAATTGCCACCGACCCGAGACGGCGCACACGATGGAG  
CTGCTCCCCGCTTCGACACCATCACCTACCGGAGCTGCTGCCGCTCGGGCGCTC  
ACCAACCGCTGGCACCGACGGGTGCGCCCCGGCACC CGTCGAATCTCGGATTC  
ACCGGCATCGACTACACCGTGTGACCTCGCCCTGATCCAGCTGGCGGGTGC  
CCGCTGAGACCGACCGCCGCCGTGGAGGCCGCTGCCCGATCGTGGCGAGACCGAACCG  
ATGCTCATGCCACCGCGTCGATCATGTCGACGCCCGCGGAACCTCGOACTACCGGC  
CACCGTCCGTCGGCAGGTGGTCTCGACCACCGGAGCAGGTCGACGACGAACCG  
GCCGTTGCCGGCGCGACGGCCCGCTGGAGACGCGGTGCCGTCGAGACACTGCCGAG  
GTGTTGCCGGCGTGGCGCCCATCTGCCCGCTGGCACCGCACGTGTTGACGAGGCCGAT  
CCACTGCGGCTGCTGATCTACACCTCCGGCAGCACCGGCCCGCAAGGGCGCATGTAT  
CCCGAGAGCAAGGTCGCCGGCATGTGGCGCGTCGGCCAAGGCCGCTGAAACAAGAT  
CAGACGGCATTCCGTCGATCACCTGAACCTCTGCCGATGAGCCACGTATGGTC  
GCCCTGCTGTGCCGGCACGCTCAGCACCGTGGCACCGCATTTGCCGCCGAGCGAT  
CTGTCGACGCTGCTGAGGACCTGCCCTGGTACGCCCACTCAGCTCGTGC  
CGGATCTGGGACATGCTCTCAGGAGTCGTCGGCGAGGTCGACCGGCCGGTGAACGAC  
GGTGCAGACCGCCCCACCGCGAGGCCGACGTGCTGCCGAACTGCCAGGAGCTGTC  
GGTGGCCGGTTCGTACCGCGATGACCGTTGCCGCCATCTCCCCGAGATGAAGACA  
TGGTGGAGACCTGCTGACATGCACTGGTCAGGGTTACGGCTGACGGAGGCCG  
GCCGGTGGTCGTCGACGGCACATCCAGGCCGCCGGTGCTCGACTACAAACTCGTC  
GCCCCGACCTCGGTAACCGACCGGCCGACCCGCCGGTGAGCTGCTGGTC  
CGCTCCACGCGACTGTTCCCCGGTACTACAAGCGTCCCACGCCGAGGTGTT  
GACGACGACGGCTCTACCGCACGGCGACATGTCGCCGAGCTGCCGCCGACAGCTG  
CAGTACCTCGACCGCGCAACACGTGCTCAAACCGCCAGGGTGAGTTCGTC  
TCCAAACTCGAGGCCGGTCTCGCCGGCAGGCCCTCGCCAGATCTCGTGTACGGC  
AACAGTGCACGCTCTACCTGTCGCCGTCGCGACGATGCCGAGCTGGCTGCC  
CACGACCCGGCATCGCTCAAGACCGGATCACGCCCTCGCTGACGCCGCGAAGACC  
GCCGGTCTGCGACTCACGAGCTGCCGCTGACTTCTCGTCGAGACTAACCGTT  
CTGGAGAACGGACTACTGACCGGCATCCGCAAGCTGGCCGCCGAAACTCAAGGCC  
TACGGCGATCGGCTCGAGGCCGCTCACGCGAACTGGCGAAGGACAGGGAGCG  
CGCACCCCTGCCGGGACGGCGGAAGCGTCCGGTGCCGAGACGGTCGGCGGCC  
GCCGCCGCTCGCTGCCGCCGCCGACGTGCGCCCCGACGCCACTTACCG  
GGCGGAGACTCGTTGTCGGCGTTGACCTCGGCAATCTGCTGAGGAGATCTCG  
GACGTCCCGGTCGGGTGATCGTCAGCCGGCGGCCGACCTCGGCGTC  
ATCGAGACCGAACAGGCTCGACCGGCAAGCGGCCGACCTACCG  
GACGCCAACAGGTACCGCGCGGACCTCACCGTGGACAAGTCA  
CTCTCCGCTGCAACAGAACTGCCGCTCCGATCGTGAGGTGCGC  
GGGCCACCGGATTCCCTGCCGCTACCTGGCGCTGGACTGGCTCG  
GTCGACGGCAAGGTACGCGCTGGTCCCGCGAAGGACGATGCC  
CTCGACGACACCTCGACAGCGCGATCCGAAAGCTGCGCA  
GCCGACCCACTCGAAGTGCCTGCCGCCGACAAGGGTGAGGCC  
CAGGTCTGGCGACTGCCGACCCGTCACCTCGTC  
AACCACGTAACGCCGTAAGCAGCTGTTGCCGCCAACCGC  
ATCCGGCTTGCCTCACCAACCGCATCAAACCGTTAC  
GGGCCGGTATCGAACCGGGCGTTCACCGAGGACGAC  
GCGGCCGGTACCGGCGACATCCGGGTGATCAGCCCG

**FIG. 8EE**

ACGGGGCCGTCGACACCGGGTACGCCAACGGCTACGGCAACAGCAAGTGGGCCGGTGAG  
 GTGTGTTGCGCGAGGCACGATCTGTGCGGGCTCCCGTGGCGGTGTTCCGGTGCGAC  
 ATGATCCTGGCCGACACCACCTACGCCGGCAGCTAACCTGCCGGACATGTTCACCCGG  
 ATGATGGTGAGCCTGGTACCGACGGGTATCGCGCCGAAGTCGTTCCACCCACTCGACGCA  
 AAGGGCCACCGCAGCGCGCCACTACGACGGGTGCGCGGTGAGTCGCTCGCCGAATCG  
 ATCTCCGCGCTGGGAGCGCAGGCAGGTGACGAGGGAAACGGTTTCCGCCACCTACAC  
 GTGATGAACCCCCACGACGACGGGATCGGCCTCGACGAATTGTCGACTGGCTCGAG  
 GCGGGGTATCGCATCGACCGCATCGACGACTACCGCGCCTGGCTGCAACGGTTCGAAACC  
 GCGCTGCGGGCGCTGCCCGAGCGCAGTCGGCAGTACTCACTGCTCCCGCTGCTGACAAC  
 TACCAGCGGCCCTGCGCACCCGATCAACGGGGCATGGCCCGACCGACCGTTCCGCGCC  
 GCGGTGCAAGGAGGCAAAGCTGGCCGGACAAGGACATTCCCCACGTACCCCCGGGGTG  
 ATCGTCAAGTACGCCACCGACCTGGAACTGCTGGGCTGATCTAG

**Amino acid sequence (SEQ ID NO:52)**

>uniprot|A3PYW9|A3PYW9\_MYCSJ Thioester reductase domain  
 MSTETREARLQQRIAHLFATDPQFAAARPDRISDAVRDDARLTAIVSA  
 VMSGYADRPALGQRAAEFATDPQTGRRTMELLPRFDITYRELLDRVRL  
 TNAWHADGVRPGDRVAILGFTGIDYTVVDLALIQLGAVAVPLQTSAAVEA  
 LRPIVAETEPMLIATGVHDVAAAELALTGHRPSQVVFDHREQVDDERD  
 AVRAATARLGDAVPVETLAEVLRGAHLPAVAPHVFDEADPLRLLIYTSG  
 STGAPKGAMYPESKVAGMWRSAKAAWNNDQTAIPSITLNFLPMSHVMGR  
 GLLCGTLSTGGTAYFAARSDSLSTLLEDLRLVRPTQLSFVPRIWDMLFQEFG  
 VGEVDRRVNDGADRPTAEADVLAEQLRQELLGGRFVTAMTGSAPISEMK  
 WVETLLDMHLVEYGSTEAGAVFVDGHIQRPPVLDYKLVDVPDFLYFSTD  
 RPHPRGEELLVRSTQLFPGYKRPDVTAEVFDDDFYRTGDIVAELGPQL  
 QYLDRRNNVLKLAQGEFVTISKLEAVFAGSALVRQIFVYGNNSARSYLLAV  
 VVPTDDAVARHPASLKTASLQQAAKTAGLQSYELPRDFLVETQPFT  
 LENGTHTGIRKLARPPLKARYGDRLEALYVELAEGQAGELRTLRRDGAKR  
 PVAETVGRAAAALLGAAAADVRPDAHTDLGGDSLSALTFGNLLQEIFGV  
 DVPGVIVSPAIDLASIAAYIETEQASTGKRPTYASVHGRDAEQVRARDL  
 TLDKFIDAETLSAATELPVPIGEVRTVLLTGATGFLGRYLALDWLERMAL  
 VDGKVICLVRRAKDDAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLAGD  
 KGEADLGLPHQVWQRLADTVDLIVDPAAVLNVHVLSQLFGPNALGTAEL  
 IRLALTTRIKPTFYVSTIGVGAGIEPGRFTEDDIRVISPTRAVDTGYAN  
 GYGNNSKWAGEVLLREAHDLCLGLPVAVFRCDMILADTTYAGQLNLPMFTR  
 MMVSLVTTGIAPKSFHPLDAKGHRQRAHYDGLPVEFVAESISALGAQAVD  
 EAGTGFATYHVMNP HDDGIGLDEFVDWLVEAGYRIDRIDDYAAWLQRFET  
 ALRALPERTRQYSLLPLLHNYQRPAHPINGAMAPTDRFRAAVQEAKLGP  
 KDIPHVTPGVIVKYATDLELLGLI

**CAR70557****Nucleotide sequence (SEQ ID NO:53)**

>gi|219932450:561817-562104 Mycobacterium leprae Br4923, complete  
 genome sequence  
 ATGAGCACCGTGGATGGTGTGTGATGAAGACATATATTACTATTATGACTGCATATATTGCGCTGCCG  
 ATGTTGCCAGCCGACTGGTATTTACCCGGAGGGACTCTTAGCTTGGCAAGCTGCTGTTATACGTGG  
 GGAGAAGCAGTATGTCTCGGCCTGAATGACTCGACGAGTGTGCGGCCGGCCGGCAACATGCCCTGG  
 CTGGACGACCCAGGCCTGCTGCTGATTGCGAGTTCCGTCGACGGCTGAATTGGCTGCTGCGCGAT  
 CCTGGTAA

**FIG. 8FF****Amino acid sequence (SEQ ID NO:54)**

>gi|219932734|emb|CAR70557.1| hypothetical protein [Mycobacterium leprae Br4923]  
MSTVDGVCDEDIYYYYDCIYCAADVVASPTGYLPGLFSLAKLLVIRGEKQYVFLNDSTVGPGPGNMPW  
LDDPGLSAVIASSVAAAELAAARSW

**YP 001220863.1****Nucleotide sequence (SEQ ID NO:55)**

>uniprot|A5CM59|A5CM59\_CLAM3 Putative acyl-CoA synthetase

ATGAGCACTGAGCAGATGGCACCGAGCAGATGGGGAGGCCAGCAGCAGGACACGTCGATC  
GAGGCATCTCGCGCAGCACGCCGACCGGACGGCGTGCAGCGATCCGGCCGGAC  
ATCACGGACATGGCTCCGGAGCTGTGGGACCGGGCGCGCTGGCCGCGCGCTC  
GGCAGACCGTGCCGCCGAGACCGCATCGCGGTCTCGGACCGGACCCGCGGACGCC  
GTCACGCTCGACCTCGCCGCGTGGATCCTCGGCCGGTGAACGTGCCCTCAGGCGAGC  
GCGCCCGTCCGCCGCGATCGAGGGAGACGACGCCGGTGTGGATCGCGGCC  
ACCGCCGACCAGGCCGCGACGCCGGGGCGTCCGCCGAGGCATCGGCGACGCCATCCGG  
ACGATGCGGCTCGACACCGACACCAGACGCCGACCGGACACCGACGCCCTGACGCTC  
GGCGCGCTCGTCGCCGCCGCGGGCGCTCGCTGCCGGAGGCCGTGGCACCCCGCACCC  
GGCGACGACCCGCTCGCGCTCCCTGCTCACACGTCCGGCAGCACGGCACGCCGAAGGGC  
GCGATGTACACCGCGCTCCATGGTCGAGCGGATGTGGCACGCCCTCGGGCCGACCCCGCC  
GCGCCCGCCGACCGCATCCACGACCGCGACGAGACGCCGAGCCATCGTCCGGTAC  
GCGTACCTCCCCATGACCCACCTCACCGCCGCTCTCCCTGCTCGCCACGCTCGGCC  
GGCGGCACGGTCGCGCTCGCACCTCGACCGACCTGTCGACGCTCTCGACGACCTGCG  
ACCTTCGCCCGACCGAGTTCGTCTTCGTGCCGCCGCTCGGGAGCTGGTGCGGCAGGAG  
GGCGACCGCGAGGAGCAGGCCGACTCACCGCCGCCAGCACGGACCGGGACGCCGGTCCGC  
GCCGAGGTCCAGGCCGACCTCGCGGCCGCCGCGTTCGGCGGACGGATCACCGGGCGATC  
TGCACCGCGACCGCTGACGCCGGAGCTCCGACGTACATCGAGGGATGCCCTCGGGCTG  
ACGTCACGACCTGTACGGGTGACGCCGGAGGCCGCGGACCTCCACGACGGGGTGTAC  
CAGCAGCCTCCCGTACCGAGACACAAGCTCGTCGACGTCCCCGAGCTGGGTACCGCACC  
ACCGACCGCCCGACCCCGCGGGCGAGCTGCTCGCAAGAGGCCCTCCGTGATCGCCGGG  
TACTTCGCCGCCGGACGTCACCGCCGCGTGTGACGAGGACGGCTCTACCGGACC  
GGCGACGTCATGGCGCAGACCGGACCCGGCACCTACGAGTACGTCGACCGTCGGAACAAAC  
GTGATCAAGCTGTCGAGGGCAGTTGTCGCGGTGGCATCGCTCGAGGGCAGTACGGC  
GGGACGCCCGAGGTGACCGAGATCGCGCTGCACGGCAGACGCCGGACGCCGTTCCCTCGTC  
GCGGTCGTCGTCGCCGGGACCCCGCGCGCTCGGAGCGCGACATCCTCGCGGCCCTCCAG  
CGCACCGCCCGGAGCACGCCCTCGCCCTACGAGGTGCCCGCGCGTGTACGTCGAA  
CCGGATCCGTTCACGGTCGACGCCGGCATGCTCTCCGACGCCGGCAAGCTCCTGCGCCTG  
CGCCTCACGCGCGGTACGGCGAGCGCCTCGCCGCCCTACGACGCCGCTCGAGGGCAG  
CAGAGCGGCACCCCTGTCGCCGCCGCTCCGCGAACGCCGCGACGACGCCGACGGTCGAC  
ACGGTGGTGCGGCCGCCCTCTGCTCTCGGGGCCGAGGTGCCCCCGCGACCGCCGCC  
GCTGCCCGGTTCTCGACCTCGCGGCCGACTCGCTGTCCGCCGCTGACGTTCTCCGGGATC  
CTCGAGGGACGTCGTCGCCGCCGAGGTGCCCGTCCGCTCCTACCGACCCGACCAACGAC  
CTCGCCGCCGTCGCCGCCCTACGTCGAGGCCGCTCCGCCCTCGACGCCGACCGTGAC  
CGCGTGCACGCCGCCGCGATCCACCCCTCCGCGTGGGCCGACCTCCGGCTCGACCGGATG  
CTCGGGGCCATCCGACGCCGGTCCCCCGGGCTCGGCCGCCGGGATCCCGCACG

**FIG. 8GG**

GTCTCTGCTCACCGGGCGCGAACGGCTACCTCGGCCGGTCTGCCATCGACTGGCTCGAG  
CGCCTCGCTCGCACGGGCGCACGCTGGTGTGCATCGCGCCGCGACGACGCCGAC  
GCCCGCGCCCGCTCGAGGCCGCGTCGCCCGCGATCCCGCGTCGCCCGCGCTTCGCC  
GAGCTGTCGGGCTCGCTCGAGGTGCTGCCCGAGACGTCAAGCGACCACGCCCTCGGCC  
GACGACGAGCGGTGGATCGACCTGCCCGCGGGTGCACCTCGTCGCGACGCCGCTGCC  
CTCGTGAACCACGTCCTCCCGTACTCGGCCGCTGTCGGTCCGAACGTGTCGTCGGCACGCC  
GAGGCCGATCCGCCCTGGCGATGCCGCCGGCAGCGTGCCCCGTACACCTCGTCTCGAGCGTC  
GCCGGTGCAGGGCGGCCGGCGCGGCCGGCGACGCCGACGCCGAGCCGTCGGCGCCCG  
GCGCTCGACGAGCACGCCGACATCCGCCGACGATCCCGAGTGGGCCGACGCCGAG  
TACGCCAACGGGTACGGCGCGAGCAAGTGGCGAGCGAGGTGTCGTCGCCGAGGCCAC  
GAGCACCAACGGCGTCCCCGTGGCGTGTCCGCTCCGACATGATCCTGGCGCACCCCCCG  
TGGCGCGGTCAAGGTGAAACCTCCCCGACGTCTTCACCCGGCTGATCTGGAGCGTGCCTCACC  
ACCGGCCCTCGCCCCCGCATCGTCGAGGCCGCGGGCCGACGGCGAGCGGCGAGCGGTGCG  
CACTACGACGGGCTGCCGCCGACTTCACGGCGGGCGATCGACGGGATCGGCCGCGCG  
CTCACCGAGGGGCCACCGCACCTCAACGTCGTAACCCCCCACGACGACGGCGTCTCGCTC  
GACACCTCGTCGACTGGATCCGCGAGGACGCCACGACATCCGCCGCGTGGACGACAC  
GCGGAGTGGGTCGACCGGTTCCGCGCGGCCGCTGGAGCGCTCCCGACGCCGACGCC  
CGGTCCGCTGCCGCTGATGACCGCTGCCCTGCCCGAGGAGCGCACGCCGCGTCTCG  
GCGATCCCGCGGAGTGCCTGCCGAGGCCGTCGCCGCGTGCGCCGCTCGGGTCCGCC  
GACATCCCGTCTCGACCAACGCCGCTCATGCCAAGGTGCGGCCGACGACCTCGCGTCTCG  
GGGCTGCTCGCGCCGGCGGGCGCGGCCGCTGA

### Amino acid sequence (SEQ ID NO:56)

>uniprot|A5CM59|A5CM59 CLAM3 Putative acyl-CoA synthetase

MSTEQMGTEQMGSQHEDTSIEAIFAQHADRTALRQRSGPDITDMGFREIWL  
DRAGALAAALGETVSAGDRIAVLGTATADAVTLIAAWILGAVSVPLQAS  
APVAALRAIVEETTPWIAATADQAATARAVAEEASGDGIRTMRLDTDTDA  
DTDTDAALTGLVARGAGLRRSPWHPPAPGDDPIALLLYTSGSTGTPKG  
AMYTRSMVERMWHALRPDPAAPADASTTADDGDAAAIVGYAYLPMSHLTG  
RSSLLATLGRGGTVVALATSTDLSLFDDLRTFAPTEFVFVPRVAELVRQE  
GDREEQRLTAGSTDRAVRAEVQADLRARAFGGRIHRAICTSAPLPEL  
RTYIEGCLGLTLHDLYGSTEAGGILHDGVIQQPPVTEHKLVDVPELGYRT  
TDRPHPRGEVVKSASVIAGYFRRPDVTAAVFDEDGFYRTGDMQAQTGPG  
TYEYVDRNNNVIKLSQGEFVAVASLEATYGGTPEVHQIALHGDSRHAFLV  
AVVVPADPAASERDILAALQRTAREHGLAPYEVPRGVIVEPDPFTVDGGM  
LSDAGKLLRLRLTQRYGERLAALYDALEEQQSGTILVAALRERADDEPTVD  
TVVRAALLLGAEVSPATAAAARFSDLGGDSLSALTFSGILEDVFGTEVP  
VGVLTDPTNDLAAVAAYVERSASDDRPTVTRVHGAGASTLRVGDRLDRM  
LGGIPTPVPRASAARPGSRVLLTGANGYLGRFLAIDWLERIAATGGTIV  
CIVRGADDADARRLEAAFAADPAFARRFAELSGSLEVLAGDVSEHRLGL  
DDERWIDLAARVDLVAHAAAALVNHVLPYSALFGPNVVGTAEEAIRLAIAG  
SVPFTVSSVAVAGGARPGATADAEPASPAGALDEHADIRATIPEWAVGDE  
YANGYGASKWASEVLLREAHEHHGPVAVFRSDMILAHPRWRQVNLPDV  
FTRLIWSVLTGGLAPASFVRRGPDGERQRSHYDGLPADFTAAAIIDGIGAA  
LTEGHRTFNVNVNPHDDGVSLDTFVDWIREDGHDIARVDDHAEWVDRFRAA  
LGALPDADRARSVPLMHAFAASPEEPHAGSAIPADAFAEAVRAVRLGPLGSP  
DIPSLDHALIAKVADDLAFGLLAPARAAAA

**FIG. HH****YP 001537947.1****Nucleotide sequence (SEQ ID NO:57)**

&gt;uniprot|A8M8D3|A8M8D3\_SALAI Thioester reductase domain

GTGACCACCAACGGAGCAGACCTTCACCGAGCGGCTCATGCCGAGGACGAGCAGATCCGG  
CGGGCCCAGGTCAAGCGCCGAAGTCTCCGCCGATGCCGGTGCAGGGCATGTCGCAGGCC  
CAGATCGTGGCCGCGATTACCGGTTACGCCGACCGCGCCGCCCTGGGTGAGCGGCC  
CGCGAAGCCGTCAACCGACCCGGTACCGGCCGACCACCCACCGGCTCTGCCATGGTTC  
GACACCATCACCTACGGCGAGGTCCGGTGCAGGGTCTGGCAGTCTCCGCCGCTGGTGG  
CACGACGTGGACGCTCCACTCCGTCCCGGCCCTCGTCGTTTCGGTCGGGTCCCCAGC  
GCCGACCTCGTGACGGTCGAGCTCGCGGTGCTACACACCGGCCGCGTCAGTGTGCCACTG  
CAGGTCAGCTCCACCGCCGAGCAACTGCCGCGATCCTCGACGAGGCCGCCCCGTCATC  
GTGGCCACGAGCGTGGACCGGCTCGCTGTGGTACCGCGATGTCGGGAACCGGTG  
GTGCGCCGGATCATGGTCCCTGAACCACGACGAGCGATCACCGCCACCAGGGATGCCGTG  
GACGCCGCGCAGTCGGCGCTGCCGGCACCGCAGTCGTCGTCACACATTGACCGAGGTG  
TTGGACCGTGGACGGGGCTGCCGCCCTGAGCCCTACCGCCGCCCCACGGGGGAGGAT  
CCCCTGTGCTGTTGATCTACACCTGGGCACTGGGTGAGCCGACAGTACCGGTACGCCAAGGGCAATGTT  
CCGGAGAGCATGACCCGCGCAACTGGGTGCGTTGCCACCCAAAGCCGACCGACATGGCG  
GTCATCCGGCTCAACTACCTGGCCTGAGCCACAAGTCGGCCGATCGTGTGTTGAG  
GCGCTCGGGTGGCGCATGCCCTCTTACCGCACACAGTACCTGTCACGCTCCTG  
GAGGACATGGCCCTGCCCGGGCCACCGAACCTGTTCTTATCCCGGGCTGCGACATG  
CTCGCCCAGCGCACGACAGCGAAGTGGCCGCGCAGTCGTCACCGCCGCGATCACGAG  
GGGGTCCGACAGGTCACACCCATCTGCGCGAGGCTGTCCTCGCCGAGGGTACCCGC  
GCGATGTCGCTGTCGCCGCGCTGAGCCACAGCTGCGTCGGTCTGGAGTCGTTCTC  
GGCTTCGCGGTGACGATGTCGGGTGACCGAGGCCGGCTGCTCGTCAACGGC  
CGGGTGTGCGCCCGCCGGTGTGACTATCGCTGGTCGACGTCGGCACCTCGGCTAC  
TTCACCACCGACCGTCCGTACCCCTCGCGGGAGCTGCTGGTGCACCGGACGATCATC  
CCCGGCTACTACCAGCGGCCGAGCTAACGCCGAGCTGTTACCGAGGACGGCTACTAC  
CGCACCGGCACATCATGGCCAGTACGCCCGACCACTCGCTATGTCGACCGCACC  
ACGAGCGTGTGAAGCTGTCACAGGGCGAGTTCGTCGACGTTGGAGGAACCTG  
TTCGCCGCTCCCGCTGATCCGGCAGATCTACCTGTAACGGCAACAGCGAGCGGCGTAC  
CTGCTGCCGTGGTCGTGCCACGGAGGAGCGCACGCCACCCGGAACCGCGGGC  
CTCAAGGGCGGTGCTGGCGAGTCGCTGCAACGCATCGCTCAGCAGCACGGCTGCACCCG  
TACGAGGTGCCGCGGACCTCCATCGAGACCAACCCGTTAGCACCACGGCTCG  
CTCTCCGACATCCGTAAGCCCTGCGTCCGAAGCTCAAGACCCGGTACGCTCCTCGACTC  
GAAGCGCTCTACACCGAGCTGCCGAGCGCGAGGCCGACCGGATCCGCACCGCTGCGGAC  
GCCGGTCCGCAACCGTGTGCCCCGGTTGCGCGAGGCTGCCGGCGTTCTCGGC  
CGCCCAGCGCAGCGCTCGACGTGAACGACCGCTTGTGGACCTCGCGGGACTCCCTG  
TCGGCCCTGGCCCTGCGAACCTGCTGAGCGACATCTCGAGGTCCGCGTCCGGTGGC  
ATCATGATCAGCGCACCGGCACCGCTCGGTTCCGTGGCGGCTGGATCGAGGCCGAGGCT  
GCCACCGCCGGAGCGGGTATCGGCCGCGACGCCACCTCCGTGACGGTGCAGACCTC  
ACCCAGGTACACGCCGATGACCTGACCCCTCGGCACGTTCTCGACGTGACGCC  
GCCGCTGCCCTGCCGCCGGCGCTGTCGACCCGGCGCTGGTGTGCTGACGGGT  
GCGACCGGCTATCTGGGCCGGTTCTGGCCCTCGAGTGGCTGGACCGCCCTTCCCGTAGC  
GGCGGGACGCTCGTGTGCGTGGTGCACGCCGCGCCGACGATGCGGAAGCCGCGGCCGCTG  
GAAAGTGTCTATGGCTCCAGCGACCCCGAGTTGCTGGAGCGCTCCGTTACTCGCCGGC  
CACGTGCGCTGTTGGCCGGCATGTTGCCGAAGCCAGGTTGCCCTGCCGGCGGGGTG  
TGGCAGGAACGGCGAAACGGTGGACCTGATCGTGCACCTCGCGGACTGGTCAACCAC  
GTTCTCCCGTACGAACAGCTGTTGGGCCAACGTCGGGGAAACGGCGGAACGGTGC  
CTCGCCGTACGGTGAAGGGAAATTGCCCTCTCCACCGTTGCCGTGATCACC  
TCGCAGACCACGACACCGACGAGGACGCGGACATCCGGCAGGGCAGGCCACCGGGTG

**FIG. 8II**

```

CTCGACGACAGCTACCGAACGGCTACCGGCCAGCAAGTGGGCAGGTGAGGTGCTCCTG
CGACGCGCCCACGAGGAGTACGGCGTGCCTCGCTCGACGTATCCTG
GCCCACTCCCGCTACGCCGGCAGCTAACGTCAGCTCCGGATATGTTACCGCCTGCTCCTG
AGCATCTGGCACCCTGATTGCCCGCAGCGTCGTTTATGCCACCGGCGACGGCGA
CGCCAGCCGGCGCACTACGACGGCTCCCGTCGACTCACCAGGGCGCTAGCCG
GTGGGTGTCACCGAGGGACACCGCACCTCAACGTAATCCACACGAGGACGGCATC
GGGCTGGATACCTTGTTGACTGGCTCGTGCAGCCGACACCCGGTGCAGCGCATCGC
GACCACGACGAGTGGGTGACCCGCTCGCAACGCCATGCGTGGCTGCCTGAACGCCAG
CGCCCGAGCTCGATCCTGCCGCTCCTACACGCCCTTGCGAGGCCGCTCGCCGACCTTC
GGATCCAGACTGCCGACGGACCCTGCGCCGCGCTGAAAGCCGCAACGTGGTCCCC
GGCAACGAGATCCCGCACCTCGATGCGGCCCTCGTACCGCAGCACCTCAGG
CTGCTCGACCTCTGA

```

**Amino acid sequence (SEQ ID NO:58)**

```

>uniprot|A8M8D3|A8M8D3_SALAI Thioester reductase domain

VTTTEQTLTERLIAEDEQIRRAQVSAEVSAAMRVPGMSQAQIVAAAGFTGY
ADRAALGERAREAVTDPTGRTTHRLLPWFDTITYGEVRSRVLAISAAWW
HDVDAPLRPGAFVVSVGVPSADLTVELAVLHTGAVSVPLQVSSTAEGQLR
PILDEAAPLIVATSVDRLAVVTAAMSGNASVRIMVNHDAITAHRDAV
DAARSALAGTAVVHTLTVLDLDRGRGLPAPEPYAAPTGEDPLSLLIYTSG
STGTPKGAMFPESTMTRANWVRFDPKPTDMAIRLNLYLPLSHNVGRIVLF
ALAVGGIAFFTAHSDSLSTLLEDMALARPTDLFLIPRLCDMLAQRHDSELA
RRRITTADHEGVRQVHHLREAVLGGRVTRAMSLSAPLSPQLRRFVESCL
GFAVHDVFGSTEAGGLLVNGRVLRRPVLDYRLVDVDPDLYFTTDRPYPRG
ELLVRTATIIPGYYQRPELNAELFTEDGYYRTGDIAMEYGPDHLYVDR
TSVLKLSQGEFVAVSRLEELFAASPLIRQIYLYGNSERPYLLAVVPTEE
AHAATREPAALKAVLGESIQRIAQQHGLHPYEVPRDLLIETPFSTANGL
LSDIRKPLRPKLKTRYAPRLEALYTELAREADRIRTLRDAGSAQPVLPA
LREAARAFLGRPAGALDVNDRFVDLGGDSLSALALSNLSDIFEVRVPVG
IMISATGTLGSVAAWIEAERATAGAGIGRATPTSVHGANLTQVHADDTL
GTFLDVTTLAAAACLPRAPLSDPRVVLITGATGYLGRFLALEWLDRLSRS
GGTLVCVVRRAADDAEAARRLESVYGSSEDPELLERFRSLAGHVRVLAGDVA
EARFGLPAGVWQELAETVDLIVHSAALVNHLVLPYEQLFGPNVAGTAELVR
LAVSVRVKGIAFLSTVAVITSQTTPDEDADIRQASPHRVLDDSYANGYA
ASKWAGEVLLRRAHEEYGVPVSVFRSDVILAHSRYAGQLNVPDMFTRLLL
SILATGIAPASFYRTGPDGERQPAHYDGLPVDFTAAAVAAVGVTGHRTF
NVLPNPHEDGIGLDTFVDWLVAAGHPVQRIADHDEWVTRFATAMRGGLPERQ
RRSSILPLLHAFAEPPPTFGSRLPTDRFRAAVKAANVPGNEIPHLDAA
LVTKYADDLRLDLL

```

**YP 001703694.1****Nucleotide sequence (SEQ ID NO:59)**

```

>uniprot|B1MCR9|B1MCR9_MYCAB Probable fatty-acid-CoA ligase FadD

ATGACCGTGACCAACGAAACCAACCCACAGCAGGAGCAGCTATCCGCCGTATTGAAAGT
CTGCGCGAAAGCGATCCGCAGTTCGGCCAGCCGCGCCAGCCGACCCGGCGTCCGAACAG
GTGCTGCGCCGGGCGCTGCATCTTCTGAAGCCATTGCGCGTTGATGACTGGATACGCT
GAGCGCCCGCGCTCGTGAGCGCGCACGCCAGTTGGTCACCGACCAGGATGCCGCACC

```

FIG. 8JJ

ACGCTGCGCTGTTGCCACGCCCTCGACACCACATACGGCAATTATGGTCCCGACA  
ACATCAGTCGCCGCTGCATGGCACCAACGACGCCGCCACCGGTTAAGGCCGGCATCTG  
GTGGCCACCCCTGGGATTCAACCAGCATCGACTACACCGTGCTGGATCTGGCATGATGTC  
CTCGGTGGCGTGGCGGTTCCGCTACAGACCAGCGCCCCGGCTTCGCAGTGGACGACCATT  
CTGGCCGAAGCGGAACCCAACACTCTTGGGTAAGCATCGAATTGATCGGCCCTGCAATG  
GAATCTGTGCGGGCACGCCCTCCATCAAGCAGGTGCTGTCGACTACACCCCCGAG  
GTCGATGATCAACGGGAGGCATTGAGGCAGCAAGCACACAACCTGCCGGCACGGCATC  
GCCATTGAGACCTCGATGCCGTATGCCCGCGCCGACTTCCGGCCACCGCTC  
TACGCACCATCGGCCGGCGACGATCCGCTGGCGCTGCTCATCTACACCTCCGGCAGCACC  
GGGGCTCCAAAGGGGCCATGCACAGCAGAAAACATCGTGCGCCGCTGGTGGATTGAG  
GACGTATGGCCGGCACCGAGAACCTGCCATGATCGGGCTGAACCTCATGCCGATGAGT  
CACATCATGGGACGCCGACCCCTACCTCCACCCCTGCTACCGGTGGAACCGGATACTC  
GCGCGTCCAGTGCACATGTCACGCTCTCGAGGACATGGAGCTGATCCGCCGACGGCC  
CTGGCCTGGTCCACGCGTGTGCGACATGGTGTTCAGCGATTCAGACCAGGGTGGAC  
CGCGCTGGCGAGCAGCAGACACCGCAGTGCAGGGCCGTTGCGGCCGAGGTCAAGGCC  
GATATCCGTGACAACCTCTTGGTGGCCGTATGGCGGTATGGTCGGTCTGCTCCG  
TTGTCCGAGGAGCTGGGTGAGTTCATCGAATCCTGCTGGAGCTGAATCTGACCGATGGC  
TACGGCTCCACCGAAGCCGGCATGGTGTCCCGCAGGGCATCGTGCACACGCCGCCGGTC  
ATTGACTACAAGCTGGTGTGACGTGCCGAACGGCTACTTCTCCACCGACAAGCCGAC  
CCGCAGCGTGAAGACCGACGGCATGTTCTGGGTACTACAAACGCC  
GAGGTGACTGCCGGCTTCGACGCCGAGGTTTACATGACCGGCACATCGTGC  
GAGCTGGCCCACGACAACATCGAGATCATCGATGCCGACACAGTGC  
CAGGAGAGTTGCGCGTGGCCACCTGGAGGCGAGTACGCCAATAGCC  
CAGGAGATCTACGCTCACGCCAGCGACGGTCTACCTGCTAGCAGTGTGGTGC  
ACGCCGGAGGCCGTTGCCGCCAAGGGCAGCGGCCACTCAAGACGACCATCGC  
GACTCGTGCAGGACATTGCAAGGAGATCCAGTGCAGTCTACGAAGTCCCCGTGAC  
TTCATCATCGAACCGCAGCATTCAACCCAGGGCAACGGCTGCTGACGGGTATGCCAAG  
CTGGCGCTCCGAACCTGAAGGCGCACTATGGACCGCGCTGGAGCAGATGTACGCC  
ATCGCCGAGCAGCAGGCTGCCAGCTTGGCGTTGACGGAGTGGACCCAGACAAGGCC  
GCGCTGGAAACGGTCTCAAGGCCGCGCAGGCCCTGCTGGCGTCTGCTGGCGAAGT  
GCCGCCAGCGCATTCAACCGATCTAGGTGGGATCTGCTGTCCGACTGTCTTCTCG  
GATCTGCTGCCGATATCTTGGCTGAAGTACCGTGGAGTCATCGTACTGCG  
AACGATCTCAGCGGTGTTGCAAATTTGATGAACAAACGCTATTGGGGGGACGCC  
CCGACCGCGAGACGGTGCACGGCGCCGGCATACGGAGATCCGGGGCGGCCACTGACC  
CTGGATAAGTTCATCGACGAGGCCACCCCTGCATGCCGACCGTGCCTCCGAAGGCC  
GGGATCCCACACCCGCTCTGTCACCGGGTCCAACGGCTACCTGGGCCACTACCTGGCA  
CTGGAATGGCTTGAGCGCCTGGACAAGACAGAAGGCAAGCTGATGCCATCGTCCGCG  
AAGAATGCCGAGGCCCTACCGCCGCTCGAGGAAGCCTCGACACCGCGACACGCC  
CTGTTGCCGACCTTCCGGTGCCTGGCGACAAGCACCTCGAAGTACTGGCCGG  
GGCAGCCCCAACCTGGCTGGATGCCGACACCTGGCAGCGCTGGCGACACCGTGC  
GTCATCGTGCACCCGCCGCCCCGGTCAACCGCTACTGCCCTACAGCCAGCTGTTGGA  
CCGAATGTCGTCGGCACCGCGAGATCATCAAGCTGCCATCACTACCAAGATCAAGGCC  
GTCACCTACCTGTCACGGTCCGGTCCGGCATATGTCGATCCGACGACATTGACGAA  
GAGTCCGATATCCGGCTCATCAGCGCGTGCCTGGACGAGCTGACGCGAACGGC  
TACGGCAACAGCAAGTGGGCGGGCGAGGTACTGCTGCCGAAAGCCCACGATCTGCGGA  
CTACCCGTGCGGTCTTCCGCTCCGACATGATCTTGGCCCACAGCGCTACACGGACAG  
CTCAACGTGCCGACCGAGTTCACCGACTAATCCTCAGCCTCATGCCACCGGAATCGCA  
CCCGCTCTTCTACCAAGCACGCCACCGCGAACGCCACTCGCCCACACTGACGGG  
CTACCCGGTGAATTACCGCCGAGGCGATCACCACTGGGACCCAGGTGGTCGACAGC  
TACGAGACCTACGACTGCCGTGAACCCGACATGCCGAGCAGGGAGTCTCGTGGACA  
GACTGGCTCATCGAAGCCGGTACCCCATGCCACGCACTGACAACACTACCGAATGGT  
ACCCGCTCGACACCGCATCGAAGCCTCCCCGAAAAACAGAAACACACTCCCTACTA  
CCACTGCTCCACGCATTGCAACAGCCGTGCCGCCGAGAACCGCGTGTCCGGCA  
AAGCGTTCCAGCACCGCTGTGCAAGGCCGCCGAATCGGTCCGGGGCAAGACGGCACT

**FIG. 8KK**

ACCGACATTCCCCACCTGTCGGCGGCTGATCGTAAATACGCCAAGGACCTCGAACAG  
CTCGGACTCCTATGA

**Amino acid sequence (SEQ ID NO:60)**

>uniprot|B1MCR9|B1MCR9\_MYCAB Probable fatty-acid-CoA ligase FadD  
  
MTVTNETNPQQEQLSRRIESLRESDPQFRAAQPDPAVAEQVLRPGLHLSE  
AIAALMTGYAERPAGERARELVTDQDGRTTLRLLPRFDTTTYGELWSRT  
TSVAAAWHDAAHPVKAGDLVATLGFTSIDYTVLDLAIMILGGAVAVPLQT  
SAPASQWTTILAEAEPTNLAVSIELIGAAMESVRATPSIKQVVVFDTPE  
VDDQREAFEAESTQLAGTGIAETLDAVIARGAALPAAPLYAPSAGDDPL  
ALLIYTSGSTGAPKGAMHSENIVRRWWIREDMAGTENLPMIGLNFMPMS  
HIMGRGTLTSTLSTGGTGYFAASSDMSTLFEDMELIRPTALALVPRVCDM  
VFQRFQTEVDRLASSDTASAEAVAAEVKADIRDNLFFGRVSAMVGSAP  
LSEELGEFIESCFELNLTGTYGSTEAGMVFRDGIVQRPPVIDYKLVDVPE  
LGYFSTDKPHPRGELLKTDGMFLGYYKRPEVTAGVFDADGFYMTGDIVA  
ELAHDNIEIIDRRNNVLKLSQGEFVAVATLEAEYANSPPVHQIYVYGSSE  
RSYLLAVVVPTEPEAVAAAKGDAALKTTIADSLQDIAKEIQQLQSYEVPRD  
FIIEPQPFTQGNGLLTGIAKLPNLKAHYGPRLEQMYAEIAEQQAAELR  
ALHGVDPKPALETVLKAAQALLGVSSAELAADAHFTDLGGDSLALSFS  
DLLRDI FAVEVPVGIVVSAANDLSGVAKFVDEQRYSGGTRPTAETVHGAG  
HTEIRAADLTLDKFIDEATLHAAPSLPKAVGIPHTVLLTGSNGYLGHYLA  
LEWLERLDKTEGKLIAIVRGKNAEAAYRRLEEAFDTGDTQLLAHFRSLAD  
KHLEVLAGDIGDPNLGLDADTWQRLADTVDVIVHPAALVNHVLPYSQLFG  
PNVVGTAEEIKLAIITTKIKPVTYLSTVAVAAYVDPTTFDEESDIRLISAV  
RPVDELYANGYGNSKWAGEVLLREAHDLCLGLPVAVFRSDMILAHSRYTGQ  
LNVPDQFTRLILSLIATGIAPGSFYQAHATGERPLAHYDGLPGDFTAEAI  
TTLGTQVVDSYETYDCVNPHADGVSLDNFVDWLIEAGYPIARIIDNYTEWF  
TRFDAIRSLPEKQKQHSLLPLLHAFEQPSAAEHGVVPAKRFQHAVQAA  
GIGPAGQDGTTDIPHLSRRLIVKYAKDLEQLGLL

**YP 001703695.1****Nucleotide sequence (SEQ ID NO:61)**

>uniprot|B1MCS0|B1MCS0\_MYCAB Probable fatty-acid-CoA ligase FadD  
  
ATGACGATCGACGCCACCGCGGACAACACCAAGGAAGCACGTCGTCAAGCGTTAGCGAC  
CGCATCAGGCCTATTCAACCGACGATGAGCAGTCCGTGCCAAGCCGATACCGCG  
GTTGATACCGCCGTCGCCAGCCTGGTCTCGCTCGCCAGGTGGTCGCCACGATCATG  
AACGGGTACCGGGACCGTCCGGCGCTCGGGCACCGAGTCCAGGAGCTCGTCGCCAGC  
GCCGGCGTTGACGCTGCGCCGTTGGCCTCCACCTGGTACACGATCCCGCCGCTCGGTACGG  
TGGGGCATGGCCCGCGTTGGCCTCCACCTGGTACACGATCCCGCCGCTCGGTACGG  
GCCGGGGACTCGTCGGATGCTCGGCTTCAACCAGCGTGGACTACACCGCTGTCGACTTG  
GCATGCATCCACCTTGGCGGGTGGCGGTTCAATTGCGAGACCGACGGCATCGGCATCCAAC  
TGGACCGCGATCCTGGCGAATCGGAACCTGGCTCGCCACGCCGTCGCTCGGCACATCACC  
GATACTGGCAATGGAATCGGTGCTCGCCACGCCGTCGCTCGGCACATCACC  
TATCATCCCGGTGTCGACGTGCAAGCGCAAAGTCTCGAATCCGCACAGCACCGGATCGCC  
GAGGCCGGCTGCCGATTTGGTAGACCCGATACCCCTGGCGATCGGCACGGGCGCG  
TTGCCGGATGCCGTTGTTCACCGCAGAGGAGGGTACCGACCCGCTGGCCCTGGTGATC  
TACACCTGGGAAGCACCGGAACCTCCAAAGGGGCCACCTATAGCGAAAAGATGGTCGCC  
AAGCCCTGGCTGCCGACACGTTGAGCTCTAAGGCCAAATTCCCTTGATCAACCTG

**FIG. 8LL**

AATTCATGCCAATGAGCCATGTGATGGGACCGCGTAGTCTGGTCACTGCCCTGGCCTGC  
GGCGGCCTGGCTACTTGCCCGTCCAGCGACATGCCACGCTGTCGAAGACATCACG  
CTTACGCCCCACCGTGGTGAACACTCGTCCCCGTGTCGACATGCTTCCAGCGC  
TACCGCAACGAGGTTGAACGCCGTACCGGGCTTGATCCGGCGGCCACCTGCCACCCCT  
GATGCCGATGTCAAGACCGATATCCCGAAGAACCTGTCGGCGGTGTCGACATGCTTCCAGCGC  
GTGTGCGGTCTGCCACTGTCCGAGGAACCTGGCCGCTTCATCGAATCCTGCCTCGAT  
GCCCGTATACCGATGGCTACGGCTCCACCGAGGCGGGCGTACCGTGCACGCTGGTACTTC  
ATTCAAGCAGCCCCCGTACCGACTACAAGCTGGTCGACGTGCCTGAGCTCGGTTACTTC  
TCCACCGACAAGCCGACCCCGCGCGAGCTGCTGTGAAAGCCGAATCGGTTCGGC  
GGCTACTCAAACGCCCCGACGTACCGCCGACGTATCGATCCGACGGGTACTACAAG  
ACCGGGGACATCGTCCCGAGCTGAGCCGACAAGATCCAGATCGTGGACCGGCGAAC  
AACGTGATAAGCTGTCAGGGTGAGTCGTGGCGATGCCAACCTGGAAGCCGAGTTC  
GCCAATAGTCCACTGGTGCATCAGATCGCTGTACGGCAGCAGCGAGCGGTGATCTG  
CTAGCGGTGGTGTGCGGACCGCTGAGGCATATGAACAAAGCGGTGGAGATGAGGATCTA  
CTCAAACGCCGTATCGGGACTCTCTCGCGCAGGTTGCCCGCAGGGCCAACCGTCC  
TACGAGGTAACCGCGCAGTCCCTGCTGGAGACCGAACCGTTCACCGTGCACGCC  
CTGACCGGCATCGCGAAGCTGGCCGACCGAACGCTCATGAGAAAGTACGGCGCCCGCTG  
GAGCAGCTGACTCCGATATCGCCGCCAGGCGCTTGAGCTGCAAGCACTGCACTCT  
GCCGGACATGAGGACAAGCCTGCTGGATACCGTGCACCGCGGTACGGCGTTGTTG  
GGGCTGTCGGCGGCCAGGTGAGCCCAGACGCCATTCTACGACCTTGGTGGCGATTCA  
CTATCCGCCCTGCCTCTGGACCTGCTGCGGATATCTTACTGTGGAGGTTCCGGTT  
GGCGATATCGTCAAGCCGCAACGATCTGACCGCTATCGCACGATCGTGGAAAGACAC  
CGGGAAAGCAGACGGTCAATTGGTAACTCCACCGCGAACCGTGCACGGTGC  
CGCAGATCCGGCCGGACCTGACGCTGGACAAGTTCATCGACCGGACACCCCTGCGC  
GCAGGCCCCGGACTGTCCACATTACCGGACCCCGCACACGGTGTGCTACCGGCGCC  
AACGGCTACCTGGGGCGTTCCCTGGCCCTGGAATGGCTTGAGCGCTGGACAAGACAGAC  
GGCAAGCTGATGCCATCGCCGGTAAGAATGCCGAGGCCCTACCGCCGCTGGAG  
GAAGCCTCGACACCGCGACACGCGACTGTTGGCGACTTCCGGTGTGGCGACAAG  
CACCTCGAAGTACTGGCCGGGATATCGGCGACCCCAACCTGGGCTGGATGCCGACACC  
TGGCAGCGCCCTGGCCGAGACCGTCGACGTACCGTGCACCCCGCCCTGGTCAACCAC  
GTACTGCCCTACAGCCAGCTGTCGGACCCATGTCGTTGGCACCGCGAACATCATCAAG  
CTGGCACTACCACCAAGATCAAGCCCATCACCTACCTTCCACAGTGGCGTGGCAATC  
TCGGTGGACCCCAAGGTATTGATGAAGACTCCGACATCCGACCGATCAGCGCGGTACGA  
CCAATCAACGACGGCTACGCCAACGGATACGGCAACCGGAAATGGCTGGCGAGGTACTG  
CTGGCCGAAGCCACCGACCTGTGCGGACTACCGTGCAGGCTTCCGCTCCGACTAATC  
CTCAGCCTCATGCCACCGGAGTCGACCCGGCTCCTCTACCAAGCACGCCACCGGC  
GAACGCCACTGCCACTACGACGGCCTGCGGATTTCACGGCATGCCATCACC  
GCCCTGGCCCATCGAGGAGTTCCACACCTACGATTGGTGAACCCGATGCCATGGG  
ATCTCGCTGGACAACCTCGTCGACTGGCTCATCGAAGCCGCTACCCCATCGCACGCATC  
GACAACACTACCGAATGGTTCACCGCTCGACACCGCCATCCAAGCCTCCCCGAAAAA  
CAGAAACAACACTCCCTACTACCAACTACACCGTGACAGGATCCACAACACCCACAC  
AACGGCGCATTCCCTGCCCGCATCAGGTTCACTGAAGCGTCCAGGCCATCTGAACGCC  
GACATCCCGCACCTCACCGGGAACTCATCGCGAAATACGCCGGACCTGAAGCAGCTC  
GGTTACTCTAG

**Amino acid sequence (SEQ ID NO:62)**

>uniprot|B1MCS0|B1MCS0\_MYCAB Probable fatty-acid-CoA ligase FadD

MTIDATADNTKEARRQRLGDRIRRLFTDDEQFRAAKPDTAVDTAVAQPGL  
RLAQVVATIMNGYADRPALEGHRVQELVADAAGRSTLRLPEFETVTYGEL  
WGMARALASTWYHDPAAPVRAGDFVAMLGFTSVDYTAVIDLACIHLGAVAV  
PLQTSASASNWTAILAESEPAVLAVSAELLDTAMESVLATPSLRHITVFD

**FIG. 8MM**

YHPGVDVQRESLESAQHRIAEGLPTISVDPPIPLAIGHGRALPDAPLFTAE  
EGTDPLALIVIYTSGSTGTPKGATYSEKVMVKPWLRACTLSSKAEIPLINL  
NFMPMSHVMGRGLVTALACGGLAYFAASSDMSTLFEDITLRTPTVVTLV  
PRVCMDMLFQRYRNEVERRTGLDPAADLATLDADVKTDIRENLFGGRVLTI  
VCGSAPLSEELAAFIESCLDARITDGYGSTEAGVIVRNNGRIQRPPVIDYK  
LVDVPELGYFSTDKPHPRGELLVKAESVFGGYFKRPDVTADVFDPDGYYK  
TGDIVAELEPDKIQIVDRRNNVIKLSQGEFVAIANLEAEFANSPLVHQIC  
VYGSRSERSYLLAVVPTAEAYEQSGGDEDLLKRLIADSLAQVAREAQLOS  
YEVRDFFILETEPFTAANGLLTGIAKLARPKLHEKYGARLEQLYSIDIAAA  
QALELQALHSAGHEDKPVLDTVQRAVTALLGLSAAEVSPDAHFIDLGGDS  
LSALAFSDLLRDIFTVEVPVGDIVSAANDLTAIARIIVERHREADGHSVTP  
TAESVHGAGHREIRAADLTLDKFIDADTLRAAPALSTFTGTPTHVLLTGA  
NGYLGRFLALEWLERLDKGLIAIVRGKNAEAAYRRLEEAFTGDTQL  
LAHFRSLADKHLEVLAGDIGDPNLGLDADTWQRLAETVDVIVHPAALVNH  
VLPYSQLFGPNVVGTAEIIKLALTTKIKPITYLSTVAVAISVDPKVFDED  
SDIRTISAVRPINDGYANGYGNAKWAGEVLLREAHDLCGLPVAVFRSDMI  
LAHSRTGQLNVPDQFTRLILSLIATGVAPGSFYQAHATGERPLAHYDGL  
PADFTASAITALGPIEFHTYDSVNPHADGISLDNFVDWLIEAGYPIARI  
DNYTEWFTRFDTAIRSLPEKQKHSLLPLHAYRHPQPHNGAFLPAIRF  
SEGVQAHLNADIPHLTRELIAKYAADLKQLGLL

**YP 001704097.1**

**Nucleotide sequence (SEQ ID NO:63)**

>uniprot|B1MDX4|B1MDX4\_MYCAB Putative fatty-acid-CoA ligase

ATGACGGCTGGTCGGCGGCTCGCGTTGCCAAACTGTTCGAGTCGATCCCCATTCCGG  
GCAGCCATGCCGGATCCAGCGGTGATGGACTCGCTGCTGGCCTGGCGCTGCCTTATCC  
CAGGTACTCCACGCCGGTGCTCACCGGTTACCGCGGACGCCCGGGTGTATGGGTTCCGGTCC  
CGCGAGTCGGTGGTCGACACCGCCACCGGGCGACGGTCGACGGCTGCTCCCTGCCTT  
GAAACCACATCACCTATGGCAACTCCTGGAAGACATCTCGGCCATCCTCGGGAGTGGCAG  
CATGGCGACATTCCCATGGCGCCGGCGACTTCATGCCACCATCGGTTCTCCAGTCCC  
GACTACGTACCCCTGGATCTGCCACCCCTCATGAATGGTCGGTCTCGATCCCAGTCAG  
CACAAACACATCTGTGGCGCAGCTGCGGATGATGCTGGAGGAGACCAGCCCACGCCCTGGT  
GCGCGAGCGCGGACTGCCTGGATCTCGGGTCGAGGCAGCTGCGGGTACCGATCTG  
CGACGGTTGTGGTGGTCGATTACCGCGCCGAGACCGACGATCATCGCAAAGGGACTGCC  
ACGGCAAGAGAACGCTTGCACGCCGGTATGGACGTTGAGTCGAACCGCTCGCAGAG  
GTGATCGGGAGAGGACGAGACCTACCCGAACCCGTGCTGTACACGGCGGGGACGATCAG  
CGCACGGCCCTGATCATGTACACCTCCGGTAGCACCGCGCGCCAAGGGGGGCGATGTT  
ACCGAGTGGACGGTACCCGCTCTGGTCTCGGGCGCCGCCAACCGGGACACCCCG  
ATCATCAACGTGAACCTCTGCCCTCAACCACCTTGCAGGGGGTAGGACTGCTGACG  
GCCTTCATCCCCGGGGCACATGCTACTTCGTCCTCCGGAGAGCGATCTGTCACCCCTGTT  
GAGGACTGGCAGCTGGCACGCCACCCATATGGGTGTGGTCCCGTGTGTCGACATG  
CTCTTCCAGCACTACCAAACGCCAGTGGACCGACTGATGGCCGGGAAACCGACGTCGAC  
ACCGCCGATCGGCTAGCCAAAACCGAAGCTGCCGAAGATGTCCTGGCGGGCGTGTGGTC  
GCCGGCATGCTGCCACCGCGCGTTGTCCCCCGAGATGAAGGCTTCTGGAGTCCCTCA  
TTGGACTTCATCTGCTTGTACGGCCTGACCGAGGTGCGCAGCGTGTCCGAGAC  
GGCAAGATTCCCGGCCGGTGTGACTACAAGCTCGTGTACGGCTGAGCTCGGG  
TACTACACCACCGACAAGCCCCATCCGCGTGGCGAATTGCTGGTCAAGAGTGCCACCGCA  
ACGCCCGGCTACTACAAGCGTCCCAGCTCACCGCCGAGGTGTCGACGCCGATGGCTAC  
TACCGCAOGGGCGATGTCATGGCGAGGTGCGGCCGACCAATTGGTGTACGTGGACAGG  
CGCAATAACGTCAAGCTGCCAGGGCGAGTCGTCGCGGCCAATTGGAAACG

**FIG. 8NN**

GTCTATGTGGGTGCGCCGCTGGTGCAGATCTTCGTCTACGGCAACAGCGAACCGCA  
 TACCTCCTCGCCGTTGTTGCCACCGAGGAAGCCTGCGGGCACACCCGACCCCGTC  
 GAACTGAAGAATTGATCCGGAGTCAGTCAGCGGACCGCCGCTCCAACCACCTGCAT  
 TCCTACGAGCTGCCGCGACTTCATTATCGAAACCACCTCATTACGATCGAGAGTGGG  
 ATGCTTGCGGCTGTCGTAAGCCGATACGTCGGAGGACTCGAGGAGCTCCGC  
 CTCGAGCAGCTACGTCGACCTCGCCGAGGCACCGTCCAGGAAGTGCAGCTCCGC  
 GATAACGGCGAACACGGCCGGCTCGATACCGTACCGAGGGCCGAGGCCCTCC  
 GGCATGTCTGGGACGCCGTCGGTCCCAGGACACCACCTCATCGACCTCGGGAGATTG  
 CTGTCGCGTTGACATTCTCCAATCTTCAGAGACTCTTCGACGTCGAGGTTCCGGT  
 GGTGTGATCACCGGCCGGCGCCGATCTGCAGCTCGCCAGCGTCAACGGCTCGACACC  
 CGGGAGCACAGCACCGCGACCGCTGCCAGCGTCAACGGCTCGACACCACCGTCATCAGC  
 GCCACCGAAGTGACACTCGACAAGTTCATCGACGCCAGACACTCCACAAACGCTTCGCAA  
 CTCGACGTGCCGGCGGGCGCGTAGCTACCGTTCTGTCACCAGGCCAACGGATATCTC  
 GGAAGATTCCCTGCCTGGAGTGGCTGCAACGGCTGCCCAGACAGGTGGACAAGTGTAC  
 TGCCCTGGTCCGCGGCGACAACGCCGATCAAGCCCTCGCGCCTCGTTGCCGCTACGGC  
 GACACCGATCGCACACTGCTCGAGGAGTCCACACCCTGGCTCGACGGCACCTGCGCTG  
 ATCGCCGCCGATATCGCTCAGCCGCGCTTCGGCGTGGATGACGCCACCTGGAGCAGCTG  
 GCGCGCGATGTCGACAAGATCGCATCCGGCCGCGTGGTCAACCACGTCGCTGCCCTAC  
 AACCAAGCTGTTGGCCCAATGTGTTGGCACGGCGAGGTTATCCGGCTGCCCTGACC  
 ACCCGGATCAAGCCGGTGACCTATCTGTCGACGATGCCGTGGCATGACCGTGCCGAT  
 TTCGACGAGGACGGGACATCCGACCGGTGAGTCCCACCCGGCATATCGACCCGGCTAC  
 GCCAACGGGTACGCCAACAGCAAATGGCCGGCGAGGTGCTGCTGGGAGGGCACACGAC  
 ATATGCGGCTGCCGGTCAGCGTGTCCCGTCCGACATGATCTGACGCCGCTTAC  
 AGCGGACAACCTCACCGTACCGACGCCCTCACCGCATGCTGAGGCTGGTCTCACC  
 GGCATCGCCCGCGAACGCTTACCAAGGGCATGGCAGCGGTGCCGCCACCGCCTCAC  
 TACGAGGGCTGCCGGTCAGCGTGTCCCGTCCGACATGATCTGACGCCGCTTAC  
 TCCGAGGGATTCGCTCGTACGACGTATGAATCCTCACGATGACGGCATTCTGTGGAC  
 ACCTTGTCGACTGGCTCATGGAAAGATGGCATTCCATCGACATCATGACAACACTACGAC  
 GAATGGCTGTCCTCGAGACGGCATTGCGAGGTCTGCCGACGAGCAGCGCGCGCC  
 TCAGTACTCCGCTCTCGATGGTATCGGATACCGGGCAACCCCGCCGTGCTGCCGCC  
 ACGCCAATCATGTATTCCGAAAGCGTACAGGAGAACACATCGGAGGTGACGGCGCC  
 GATATTCCGAAATCGATCGTGCCTGATGCCAAATACATGCCGATCTACGAGCACAC  
 AGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:64)**

>uniprot|B1MDX4|B1MDX4\_MYCAB Putative fatty-acid-CoA ligase

MTAGAAARVAKLFESDPQFRAAMPDPAVMDSLLAPGLRLSQVLHALLSGY  
 AERPVMPGRSRESVVDTATGRTVDRLLPAFETITYGQLLEDISAILAEWQ  
 HGDIPMGAGDFIATIGFSSPDYVTLDLTLATLMNGSVSIPLQHNTSVAQLRM  
 MLEETSPLRVAASADCLDLAVEAAVGLTDLRRVVVDYRAETDDHREKLA  
 TARERLHAAGMDVVEPLAEVIGRGRDLPPEPVLYTAGDDQRTALIMYTSG  
 STGAPKGAMFTEWTVTRFWSSGAAPNRDTPIINVNFLPLNHLAGRVGLLT  
 AFIPGGTCYFPESDLSTLFEDWQLARPTHMGVVPRVDMLFQHYQTRVD  
 ALMAGGTDVDADRIAKTELREDVLGGRVVAAGMLATAPLSPEMKALESS  
 LDFHLLDLYGLTEVGGVFRDGKISRPPVLDYKLVDVPELGYYTTDKPHPR  
 GELLVKSATATPGYYKRPDVTAEVFDADGYYRTGVDVMAEVAPDQLVYVDR  
 RNNVIKLAQGEFVAVANLETVYVGAPLVRQIFVYGNSERAYLLAVVPT  
 EALRAHPDPVELKNSIRESLQRTARSNHLHSYELPADFIIEETTPFTIESG  
 MLAADVKGPIRPMIEHYGDRLEQLYVDLAEARVQELRQLRDTAQQRPVLD  
 TVTEAAQALLGMSADAVRPDHFFIDLGGSDSLALTFSNLLRDLFDVEVPV  
 GVITGPAADLRKLAAYIQHEREHSTATASAHSVHGLDTVISATELTLDFKI  
 DAETLHNASQDVPAGAVATVLLTGANGYLGRLCLEWLQRLSQTGGQLI

**FIG. 800**

CLVRGDNAADQALARLVAAYGDTDRLLLEFFTLARRHLRVIADIAQPRF  
GVDDATWEQLARDVDKIVHPAALVNHVLVPNQLFGPNVFGTAEVIRLALT  
TRIKPVTYLSTMAVAMTVPDFEDGDIRTVSPTRHIDPGYANGYANSKWA  
GEVLLREAHDICGLPVSFRSDMILTHRRYSGQLNVTDAFTRMLLSLVLT  
GIAPRSFYQGDGSGARPRAHYEGLPVDFVTEAITSLGLSSSEGFRSYDVM  
NPHDDGISVDTFVDWLMEGDHSIDIIDNYDEWLSRFETALRGLPDEQRRA  
SVLPLLDAYRIPGNPRRAATPNHVFRKAVQENNIGGDADIPOIDRALI  
AKYIADLRAHRL

**YP 001705436.1**

**Nucleotide sequence (SEQ ID NO:65)**

>uniprot|B1MLD7|B1MLD7\_MYCAB Probable fatty-acid-coa ligase FadD  
  
ATGACTGAAACGATCTCCACAGCGGCTGTCCCCACTACGGATCTCGAACAGAGCAGGTGAAG  
CGACGCATCGAGCAGGTGTCCAACGATCCGAGCTGGCGCGCTTCTCCCGGAAGAT  
TCGGTCACCGAGGCGGTCAACGAGCCCGATCTACCGCTGGTGAGGTGATCAGGCGACTG  
CTGGAGGGCTACGGTGACCGCCACTCGGCCAGCGCGCTTCGAGTTCTGACCGGG  
GACGACGGTGCACCGTGATCGCCTGAAGCCCGAATACACCACCGTCTCCTACCGCGAG  
TTGTGGGAACGTGCCGAGGCTATCGCTGCCGTGGCACGAGCAGGGCATCCGTGACGGC  
GACTTCGTCGCTCAGTTGGTTTACCAAGCACGGACTTCGCGTCGCTCGACGTGCGGGGA  
TTGCGTCTGGCACCGTCTCGGTGCCCTGCAGACGGGCGCTCGTGCAGCAGCGCAAC  
GCGATTCTCGAAGAGACCCGGCCCGAGCTTGCCTGCGAGTATCGAATACCTTGATGCC  
GCCGTCGATTGGTGTGCGACCCCTCGGTGCGACTCCTCTCGGTTTCGACTATCAC  
GCGGAGGTGACAGCCAGCGAGCGCTGGAGGCTGTGCGGGCCCGGCTTGAGAGTGC  
GGCGGAGCATCGTCGAGGCCCTGGCGAGGCTCTCGCGCGGGGACCTGCC  
GCCGCGCCGCTGCCAGTCAGATCCCAGTCGCTCATCTACACCTCCGGC  
AGCACCGGTACCCCAAGGGCGCCATGTATCCGAATGGCTGGTGCACACTGTGGCAG  
AAGAAGTGGCTACCGACGATGTATTCCGTCATAGGGCGTAACCTCATGCCCATGAGC  
CACCTGGCGGGTGCCTCACTCTCATGGGACCCCTTCCGGTGGCGAACCGCCTACTAC  
ATCGCTTCGAGCGATTTGACTTCTCGAGGACATCGCCTCATCCGCCCTCCGAA  
GTGCTCTCGTGCCGCGTGTGGTGGAGATGGTGTCCAGCGTTTCAGGCAGAATTGGAC  
CGGTCCCTGCCCCGGGTGAGAGCAACTCCGAGATCGGGAGCGAATCAAGGTCCGCATC  
CGGGAACAGGACTTCGGCGGTGTGCTAGTGCTGGCTCCGGTGGCGCCCGTTGTCT  
CCTGAGATGACGGAGTTCATGGAGTCGCTGTCAGGTGCCGTGCGACGGGTATGG  
TCCACCGAGGCCGGTGGTGTGGCGTACGGAGTCCTGCAGCGTCCGCCGTACCGAC  
TACAAGCTGGTGACGTTCCGAACTCGGATACTTCACACAGATCGCCGATCCCCGT  
GGCGAGCTGGGTTGAAGTCGGAGACGATGTTCCCGGCTACTACAAGCGCCGGAGACC  
ACTGCCGATGTCTCGATGACGAGGGGTACTACAAGACGGTGACGTGGTCGCCGAGCTC  
GGCCGGATCACCTCAAGTACCTCGACCGCGTCAAGAACGTCCTCAAGCTCGCGAGGG  
GAGTTTGTGGCTGTCAGGCTGGAGGCCCTAACCGGAGCGCCGCTGGTCCGGCAG  
ATCTTGTGTACGGGAACAGTGAACGCTCGTCCCTGCTGGCTGTGTTGGCAGACACCC  
GAAGTCCTTGAGCGGTACGAGATTGCCAGATGCGCTCAAGCCCTGATCCAGGATTG  
CTGCAGCAGGTGCCAAGGACGCCAGCTGCAATCCTATGAGATACCGCGACTTCATC  
GTGAGACGGTGCCTCACCGTCAGTCCGGATTGCTATCGGACGCGCAAAGCTGCTG  
CGCCCGAAGCTGAAGGATCACTACGGAGAGAGGCTGGAGGGCGCTGTACGCCGAACTGGCG  
GAAAGCCAGAATGAGCGGCTGCCAGTTGGCCAGGGAGGCAGCCACGCCCGGTCTG  
GAGACGGTGACCGATGCCGCCGCCGCTGCTGGCGCATCGTCCTCGGATCTGGCTCCT  
GATGTGCGATTCATCGACCTCGGTGGCGACTCACTGTGGCGCTGTGACTCCGAGCTG  
CTGCAGCAGATCTTGAGGGTGGACGTTCCGGTGGCGTCAACAGCGTCGCCAACGAC  
CTTGCGCGATGCCGCCACATCGAGGCCAGCGGACGGCGCCGCTACGCAGCCGACC  
TTTGCCTGGTCCACGGCAAGGACGCCAGGGTCACTACCGCCGGTGAACTCACCCTCGAC

**FIG. 8PP**

AAGTTCTTGGACGAGTCACTGTTGAAAGCGGCCAAGGACGTTCAAGCCGGCAACGGCCGAT  
GTCAGAACCGCTTAGTGACCGGGCAACGGCTGGTGGGTGCTGGCTGGTGCATCGAT  
TGGCTGGAGCGGTTGGCACCCAAATGGTGGCAAGGTCTACGCCCTCATTGCGCCGCGAT  
GCCGAAGCAGCCCAGGGCACGGTGGACGCCGTACGAATCGGGTGAATCCAAGCTGTCC  
GCGCATTATCGTCAGCTGGCGCAACAGAGTCTGGAAGTTATGCCGGCGATTTCGGCGAC  
CAGGATCTCGGTCTATCCAGGAAGTTGGCAGAAGCTGGCCAAGGACGTGGACCTGATC  
GTGCACTCCGGTGCCTGGTGAACCACGTGCTGCCGTACAGCCAGTGTGTCGGTCCGAAT  
GTGGCGGGTACCGCCGAGATCATCAAGCTGGCAATTCCGGAGCAGTCAAGCCGGTCACC  
TACCTGTCGACGGTGGCATGCCGACCAGATTCCGGTGACGGAGTCGAGGAAGACTCC  
GATGTTCTGTTGATGTCGGCCGAGCGCCAGATCAATGACGGCTACCGAAGCGGATACGCC  
AACTCAAAATGGGCCGGCGAGGTGCTGTTGCGGGAGGCTCATGACCTAGCGGGGCTGCC  
GTGCGTGTGTTCCGCTCGACATGATCCTGGCGCACAGTGAACGACTACCGACAGCTCAAC  
GTCACCGACGTGTTCACCCGGAGCATCCAGAGTCTGCTGTCACCAGTGTGCAACCGGCC  
AGCTTCTATGAATTGGATGCCGACGGCAATCGCAGCCGCTCACATGACGGTGTGCC  
GGCGATTCACCGCCGATCGATACCGCCATCGCGGTGTGAACGTGGTAGACGGTTAC  
CCAGCTTCGACGTGTTCAACCCGACCATGACGGTGTCTCGATGGATACTTCGTCGAC  
TGGCTGATCGACGCAAGCTACAAGATCGCGGGATCGACGATTACGACAGTGGCTGCC  
CGGTTCTGAGCTGGCCTCAAGGGATTGCCGAGCAGCAGCGCAACAGTCGGTGTGCA  
CTTCTCAAGATGTACGAGAAGCCCAACCGGCATCGACGGAAGTGCACCTCCGACCGCA  
GAATTCAAGTCGCGCCGTGCACGAGGCGAAGGTCGGAGACAGCGGTGAGATAACCGCACGTC  
ACCAAGGAGCTGATCCTCAAGTACGCCAGCGATATTCAAGCTGTTGGCCTGGTAG

**Amino acid sequence (SEQ ID NO:66)**

>uniprot|B1MLD7|B1MLD7\_MYCAB Probable fatty-acid-coa ligase FadD

MTETISTAAVPTTDLEEQVKRRIEQVVSNDPQLAALLPEDSVTEAVNEPD  
LPLVEVIRRLLEGYGDGPALGQRAFEVTGDDGATVIALKPEYTTVSYRE  
LWERAEAIAAAWHEQGIRDGFVAQLGFTSTDFASLDVAGLRLGTVSPVPL  
QTGASLQQRNAILEETRPAVFAASIEYLDAAVDVSLATPSVRLLSVFDYH  
AEVDSQREALEAVRARLESAGRTIVVEALAEALARGRDLPAAPLPSADPD  
ALRLLIYTSGSTGTPKGAMYPOQWLVANLWQKKWLDDVIPSIGVNFMPPMS  
HLAGRLLTMGTLGGGTAYYIASSDLSTFFEDIALIRPSEVLFPVRVEM  
VFQRFQAEELDRSLAPGESNSEIAERIKVIREQDFGGRVLSAGSGSAPLS  
PEMTEFMESLLQVPLRDGYGSTEAGGVWRDGVLQRPPVTDYKLVDVPELG  
YFTTDSPHPRGELRLKSETMFPGYYKRPETTADVFDDDEGYYKTGDVVAEL  
GPDHLKYLDRVKNVLKLAQGEFVAVSKLEAAYTGSPLVRQIFVYGNERS  
FLLAVVVPTPEVLERYADSPDALPKLIQDSLQQVAKDAELQSYEIPRFI  
VETVPFTVESGLLSDARKLLRPKLKDHYGERLEALYAEALAESQNERLRQL  
AREAATRVLETVDAAAALLGASSSDLAPDVRFIDLGGDSLSALSYSEL  
LRDIFEVDVPGVINSVANDLAIAARHIEAQRTGAATQPTFASVHGKDAT  
VITAGELTLDKFLDESLLKAADKVQPATADVKTVLVTGGNGWLGRWLVL  
WLERLAPNGGKVYALIRGADAEEAARARLDAVYESGDPKLSAHYRQLAQQS  
LEVIAGDFGDQDLGLSQEVWQKLAQDVLDLIVHSGALVNHLPSQLFGPN  
VAGTAEIIKLAISERLKPVTYLSTVGIADQIPVTEFEEDSDVRVMSAERQ  
INDGYANGYGNNSKWAGEVLLREAHDLAGLPVRVFRSDMILAHSDYHGQLN  
VTDVFTRSIQSLLLTGVAPASFYELDADGNRQRAHYDGVPGDFTAASITA  
IGGVNVVDGYRSFDVFNPVHDGVSMDFVWDWLIDAGYKIAIRIDDYDQWLA  
RFELALKGLPEQQRQQSVLPPLLKMYEKPQPAIDGSALPTAEFSRAVHEAK  
VGDSGEIPHVTKELILKYASDIQLLGLV

FIG. 8QQYP 001828302.1**Nucleotide sequence (SEQ ID NO:67)**

&gt;uniprot|B1VMZ4|B1VMZ4\_STRGG Putative carboxylic acid reductase

ATGGCCGAACCGCTGGACGCCAACCGCGTCGGCGCACGATCCGGGCCAGGGGCTGCC  
GAGGCCCTGGCCGCCGTCGAACCGGGCCGGCGCTCGCGAGGTCACTGGCTCCGCTG  
GAGGGCCACGGGACCGGCCCTCGCGAACCGGGCCGGAGCGGAGACCGGGCGT  
CTCCTCCCGCACTTCGACACCACAGCTACCGCGAACGTGGTCCCGTCCGTGCGCTG  
GCCGGCGGTGGCACCGACCGACCCGAATAACCCCTGGGCCCGCGACCGGATCTGCACC  
CTCGGCTCACCAAGCACCGACTACCGCGACGCTGGCGTGCATCCACCTGGGGCC  
GTGCCCGTCCCCCTCCGTCACGCCCGTGCCTGGCGGACTGGCGCCGGTGTGAGGAG  
TCCGGCCGACGGTACTGGCGAGCGTCGACCGGCTCGACACCGCATCGACGTGTC  
CTCGCGTCGAGCACGATCCGCCGCTCTCGTCTCGACGACGGCCGGGGCCACCGC  
CCGGCGGGGCACTGGCGGCCGCAACGCGTGTCCGGCAGCCGGTACCGTCGAC  
ACCCCTGGCCGGACTCATCGACCGGGCAGGGACCTGCCGCCCCCGCCCTGTACATCCG  
GACCCGGGGGAGGACCGCTGCCCTGCTCATCTACACCTCCGGCAGCACGGCGGCC  
AAGGGCGCATGTACACCCAAACGGCTCCTGGCACCGCGTGGTACGGGTTAGCTACGG  
GCGGCCGACACCCCGCATCGCGTCTACCTGCCGAGAGCCACCTCGCGGGCCGC  
TACCGGGTGTGGATGGATCGCTCGTAAGGGGGCACCGCTACTTCACCGCTGCCGACGAC  
CTGTCACCCCTGTCGAGGACATGCCCTGGTCCGCCACGGAGCTGACCATGGTCCCG  
CGCCCTGTGCCGACATCGCTCCAGCACTACCGGAGCGAACGGGACCGCCGGCGACGAA  
CCGGCGACATCGAGGGCGGTACGAAGGGGGTGCAGGGGACTTCCTGGCGGGCGC  
GTCGCCAAGGCCTCGTCGGCACCGCGCTCTCCGGCAACTCACGGCGTCGAG  
TCCGTCTCGGCTTCCACCTCTACACCGCTACGGCTCCACCGAAGCCGGCGAGTGCTG  
CTGGACACGGTGGTGCAGGCCCTCCGGTACCGACTACAAACTGGTCGACGTCCCCGAA  
CTGGGCTACTACCGGACCGACCTGCCCATCCGCGCGCACTGCTGCTGAAGTCCCAC  
ACGCTCATCCCCGGACTACCGGCGCCGACCTCACCGCGCATCTCGACGCGGAC  
GGCTACTACCGCACCGGTACGTTCTCGCGAGACCGGACCCGACCGGCTGGTACGTC  
GACCGCACGAAGGACACCCCTGAAGCTGTCCAGGGCGAGTTCTGGCGTGTCCGCTG  
GAGACCGTCTCTCGACAGCCCTCTCGTCCAGCACCTCTACCTGTACGGCAACAGCGAG  
CGCGCCTACCTCTCGGGTGGTGGTGCCTACCCGGACCGCGCTGGCGGGTGCAGGG  
GACACCGAGGGCGTCAAGGGCGTGTGCTCATGGAATCCCTCCGACGGTGTGCG  
GGGCTCAACCGTACGAGATCCCGCGGGCATCTCGTCAAGCCCTCAGCCCG  
GAGAACGGCTCTTCACCGAGAGGCCACAAACTGCTGCCGCCCCGCCCTCAAGGAGCGCTAC  
GGGCCCGCTCTGGAGCTGCTGACCGACTGCCGACGGGAGAGCCGGCTGCG  
GAGCTGCCGCCGACCGGTGGCGACCGGGCGTGCAGGAGACGGTCTCCGGGGCGCCAG  
GCCCTGCTGGGATCCCCGGCTCCGACCTCCGGCCGGCGCAGCTACCGACCTCGG  
GGGGACTCCCTCTCCGGCTCTCGTTCTCGAGCTGATGAAGGAGATCTTCCACGTCGAC  
GTCCCGGTGGTGCATCGGCCGGCGACCTGGCGAGGTGGCGGGTACATC  
ACGGCGGCCGTCGGCGGGGACCGCGGCCACGCCCGCTCCGTGACGGGGAG  
CACCGCACCGAAGTCCGCGCCGGGACCTCGCCCCGGAGAACGTTCCCTCGACGCGCCACG  
CTCGCCGCCGCTCCGGCGTCCCCGCCGACCGCGACGTCCGGACGGTCTGCTGACC  
GGGCCACCGCTACCTCGGCCGTTCCCTCGCTGGAGTGGCTGGAGCGGCTGGCGCC  
TCGGCGGACGGCTGGTCTGCCCTCGCGGACCGCGACGCCCGCTCCGTGACGGCG  
CTGGAGGCAGCCTCGACAGCGCGACACGCCCTGCTCCGGCGCTACCGGAAGGC  
GGGAAGACCTGGACGTGGTGCAGGGGACATCGCGAACCAACTGCTGGGCTGGCGAG  
GAGACCTGGCGGGAACTGGCCGGCGCGTGGACCTGATCGTGCACCCGGCGCGTGGTC  
AACACACTGCTGCCGTACGGCGAACGTGGCCCAAACGTCGTCGGCACCGCCGAAGCG  
ATCCGGCTGGCGCTACCCACCGCTGAAGGGCGTCAACCACGTCGACCGTGC  
TGCCTCGGACCCCCGGCAGAGACGGCGACGAGAACGGCGACATCCGGCGCGTCCCG  
GTACGGACCACCGGCCAGGGTACGCCGACGGATAACCGGACCGAACGAAATGGGCCGGCGAG

**FIG. 8RR**

GTCCTCCTGCGTGAGGCACAGAGCGCTATGGCTCCCGTGCCTCGTCCGGTCCGAC  
 ATGGTCCTGGCGCACCGCACCTACACCGACAGGTCAACGTCCCCGACGTCTCACCCGG  
 CTGCTGCTCAGCCTGGTCGCCACCAGCATCGCCCCCGCTCGTTCTACCGCACGGACACC  
 CGTCCCCACTACGACGGCTGCCGGTCGACTTCACCGCGGAGGCCGTCGTCGCGCTGGC  
 GCACCGATCACCGAGGGCCACCGGACCTCAACGTCTCAACCCGACGACGACGGCGTT  
 TCCCTGGACACCTTCGTCGACTGGCTCATCGAGGCCGCCACCGATCCGGCGGATCGAC  
 GACCACGGTGCCTGGCTACCCGCTTCACCGCGCGCTCCGCGCCCTGCCGGAGAAGCAG  
 CGGCAGCAGCTCCCTGCTCCGCTGATCGGCGCTGGCGGAACCCGGCGAAGGAGCCCCC  
 GGGCGCTGCTCCCCCGCCGGCTTCCACGCCGCCGTCGGGGGGGGGTCGGCCCC  
 GAGCGGGACATTCCCCGGGTGTCGCCGGACCTCATCCGCAAGTACGTCACCGACCTGCGC  
 GCACTCGGCTCCTCGCCGGCCCCCTGA

**Amino acid sequence (SEQ ID NO:68)**

>uniprot|B1VMZ4|B1VMZ4\_STRGG Putative carboxylic acid reductase

MAEPLDAATASAHDPGQGLAEALAAVEPGRALAEVMASVLEGHGDRPALG  
 ERAREPETGRLLPHFDTSYRELWSRVRALAGRWHHDPEYPLGPGRIC  
 LGFTSTDYATLDLACIHLGAVPVPLPSNAPIPRLAPVVEESGPTVLAASV  
 DRLDTAIDVVLASSTIRRLVFDDPGPATRPGGALAARQRLSGSPVTVD  
 TLAGLIDRGRDLPPPPLYIPDPGEDPLALLIYTSGSTGAPKGAMYTQRLL  
 GTAWYGSYGAADTPAISVLYLPQSHLAGRYAVMGSVLVKGGTGYFTAADD  
 LSTLFEDIALVRPELTMVPRLCMDMLLQHYRSERDRADEPGDIEAAVTK  
 AVREDFLGGRVAKAFVGTAPLSAELTAFVESVLGFHTGYGSTEAGGVL  
 LDTVVQRPPVTDYKLVDVPELGYYATDLPHPRGELLKSHTLIPGYYRRP  
 DLTAEIFDADGYYRTGDFVAFETGPDRLVYVDRTKDTLKLSQGEFVAVSRL  
 ETVLLDSPLVQHLYLYGNSERAYLLAVVVPTPDALAGCGGDTEALRPLL  
 ESLRSVARRAGLNAYEIPRGILVEPEPFSPENGLFTESHKLLRPLKERY  
 GPALELLYDRLADGQDRLRELRRTGADRPVQETVLRRAQALLGSPGSDL  
 RPGAHFTDLGGDSLSAVSFSELMKEIFHVDPVGAIIGPAADLAEVARYI  
 TAARRPAGAPRPTPASVHGEHRTEVRAGDLAPEKFIDAPTIAAAPALPRP  
 DGDVRTVLLTGATGYLGRFLCLEWLERLAPS GGRLVCLVRGSDATVAARR  
 LEAAFDSGD TALLRRYRKAAKGTL DVVAGDIGEPLLGLAEETWRELAGAV  
 DLIVHPAALVNHLLPYGELFGPNVVGTAEARLALTTRLKPVNHVSTVAV  
 CLGTPAETA DENADIRA AVPVRTGQGYADGYATSKWAGEVLLREAH ERY  
 GLPVAVFRSDMVLAHRTYTQVNVPDVLT RLLSLVATGIAPGSFYRTDT  
 RAHYDGLPVDFTAEAVVALGAPITTEGHRTFNVLPNPHDDGVSLDTFVDWL  
 EA GHPIRRI DDHGAWL TRFTA ALRALPEKQRQHSLLPLIGAWAE PGE GAP  
 GPLLPARRFHAAVRAAGVGPERDI PRVSPD LIRKYVTDLRALGLLAGP

**YP 001851230.1****Nucleotide sequence (SEQ ID NO:69)**

>uniprot|B2HE95|B2HE95\_MYCMM Fatty-acid-CoA ligase FadD9\_1

TTGTCAATTACCTGTGGATACCGTGACAGCGGAGCGCCCGTCGCATCGAGCAGCTT  
 TACTCCACCGATGCGCAATTGCCGCCGCCAGTACGGCGGTGGTATCGCAATC  
 AGCAAGTCCGGTTGGGATTACACAGATCATTCAAACGGTATGGACGGATACCCGCAA  
 CGTCCGGCACTTGGCAGCGGGAGCGCGCTTACCGATCCAATACGGCGTAGC  
 TCGCGCAGCTGTTGGCAGTTGAGACCACACCTACCGGAGTTGTGGAACCGCACC  
 AATGCATTGACCAACGCATTGCCGCCAGGCACCTGGGATCGGGTCAAGGGGTCTGT  
 GTGCTGGGATTCGCGAGCATCGACTACGCCACCATCGACTTGGCGCTGATGTTGCTCGGC

FIG. 8SS

GC GG TAT CG GTT CC GGT GCG ACG AAT GCG GCT CG CCCC AGC TGT GCC AT AT CG CT CC  
GAG ACC CAG CCC AG CCT GAT CG CT CG AG TAC GAA AAC CT GCC C GAT GCA AT CT CT TT G  
GT GCT GT CG CACC CG CG ACC AC ACC GGG TGG TGG TCG ACT ACC GCCC GA CT CG AC  
GC AC ACC CGA AGC CCT CGA AG CG CT CG CG CG CT GG CC CAT CCC GT GAC CG TC  
GAA AC GCT CACC CG CAT CAT CG CG CG CG CGA AC CG GT GCG CG CG GAG GCG ATT TG C  
GG CG CC CAG TCC CG CT GAT GC ACC CG CG CT TT GAT CT AT AC CT CG GA AG CAC CG GG CA  
CCA AGG CGT CGT CT AC ACC CGA ACC CG GT GG CG ACT TCT GG CG CACT CG AA AG CC  
GAG GT CGA AG CG ACC GA AAC AGA ACC CG CT CC TG AT CAC CT CA ACT TCT AT CG CG AT G  
AG CC AC CG GA AC CG CC CAG GT GCT CT AC CG GAC GCG CT GT CC AAC CG CG ACC CG CG AT  
TT CAC CG CC CG CAG CG AC CT CT CG AC GCG CT TG AT GAT CT CG CG TT GG TCC CG CC ACC  
GA AT TGG GCT TT CC ACC CG CG CATT GG GAC AT GCT GT GG AG AG GGT TT GG CG CG AAG TC  
GAC CGT CG GCT CC GG AC CG CG AC AG CG GAG GG CG CC ACC CG GG CG CT GA AGG CT CG  
GT GG CG CC GAC CT AC CG CAG GT GCT CG CG AC CG GT AT CG CG CT GG CG AT GAT GG G C  
TCC CG CC AAT CT CG GAG CAG AT GAA AG C AT CG TG AAT CC CT TG CT CG AT CT GG AC GTC  
AT GG AG GG CT AT GG CT CC AC GG AAG CG GG A AC GG T CAT CAT CA AC A CG AG GT 'TC AG CG  
CCCC AG GT GAT CG ACT ACA AG CT GG TG AC GCG T CG CG A ACT GG GCT AT TT CC TT ACC GAC  
CG GG C AT AT CG CG GG GG CG A ACT GCT GG T CAAA AC CG GG AC ACT GT TT CC GG CT ACT AC  
CG GG ACC CC GA AG AC CG CG CC CAG GT CT TG AC CC CG AC GG CT CT AC CG GG AC CG GG AC  
AT CAT GG CC CA AG TO GG CC CG AT CG GCT CG CT AC CT CG AC CG CG CG CA AC A CG T GCT G  
AAG CT GT CG CAG GGG GAG TT CG CG GT CT CG GACT AGA AG CA AT AT TG CC A AT AG C  
CC GT TGG TCC CG G CAG AT CT CG CT AT GCA AC CG GT GCT CG CG CT ACC C ACT GG CG TA  
GT CG TCC C ACC CAG GAC GCG AC AG T CG CG CC AC CG GT CG CG CG A ACT CA AGG CG A ACT C  
CAT AC AT CG CT CG ACC CG GT TG CC AT GT CG CC CG GT CT GG C ACC CT AC GAG AT CCC AC CG  
GACT T CAT GT CG AG A CA ACC C C TT CAC CG CG CAG A AC CG CG CT GCT C ACC CG A AT CC AC  
AAG CT GG CC CG CG CAC CT CAC G CAG CG CT AT GG CG CAC GT CT GG AG GCT GT GCT GAC C  
GAG CT GG CC GAC AG CG CAG ACC CG CG CT G CAC CG ATT CG CG CC AA ACC CG GT GG CG GG CT G  
CC CG CG CT CG AG ACC AT CAG CG GT GCG C CG GG ACT GT GG CAC GG AG ACC ACC CG AG  
CC CG CG CC CG AG GCG CACT T CAG CG CAG CT CG CG GT TG CG CG GT AT CG CG GT AC GT TC  
TCC AAC CT GCT AC AG CAC AT CT AC GG TT CG AT GT TT CG GT CG GT AT CG CT CG CC CG  
GCA ACC GAT TT CG GG CG CT GG CC AG CC AC GT CG AG AG CG CG CG GT CG CG GT GG GAT GGT CG  
GGG C CAG CT CG CG TG CG AC GT G C C C G G G G G A C C T C G G T A C C G C G G G A C C T G  
AA ACT TG G C A A G T T C T G G A C A C C A A G A C A C T C G C A G C T G C C A C G G C T G C C G T G C C  
GAT G C C C G G G C A C G G A C G G T G C T A C T C A C C G G C G A C C G G A T T C C T G G G A C G C T A C C T G  
G T G C T G G A T T G G C T G C C C G G T T G C G G G C G T C G G C G G C A A G C T G A T C T G T C T G G T C G C  
G C C G C G T C C G A C G A A C A A G C C C G G T T C G G C T G G A T A C G G C T T C G A T A G C G G C A T C C G  
C A G C T G C C C G A G C A C T T C G G C A G C T C G C G C T G G A G G T C C T C G C C G G C A T  
A A G A G C G A A C C A G G T C T C G G T C T G G A C G G C C C A A C C T G G C A G C G A C T G G C C G A C A C G G T C  
G A C C T G A T C G T C G A C C C C G C C A C G C T G G T C A A C C A C G T G C T G C T A C C G G C A G C T G T C  
G C T C C C A A C G T G G C G G C A C C G C G A G G T T G C T C C G C C T C G C A C T C A C C A C C A A C G C A A G  
C C C T A T G C C T A C G T C T C G A C C G T C A G C T G G C G C A C C A G A T C G A A C C G T C C G C A T T C A C C  
G A A G A C G C C G A C A T C C G G G A G A T C A G C C G C A C C C G A A C C A T C G A T G A C A G C T T G C C A A C  
G G C T A C A C C A C A G C A A G T G G G C A C G C A G G G T G C T G T G C G T G A G G G T C A C G A T C T G T G C  
G G A C T G C C G G T C A C G G T C T T C G T T G C G A C A T G A T C T G G C G G A C A C C A G C A G C T A C G C C G G C  
C A G C T C A A C C T C G C C G A T A C C T C A C C C G G T G A T G C T C A G T G T G G C G G C A C C G G G A T C  
G C G C C C G C C T C G T T C A C C G G T G G G C C C G A C G G C A A A C G C C A G C C C G C C C A C T T C G A C  
G G A T T G C C C G T C G A A T T C A T C G C C G A G G G C G T G G C C A C C C T G G G G G C G G C G C C A C G A C  
G G G T T C C A G G T C C A C C A T G T G G C G A A T C C G C A C C A C G A C G G C G T T G G G T T G G A C G A G T A C  
G T C G A C T G G C T A G T C G A T G C C G G T G C C C A T C C G G C G A T T C C C G A C T A T G A C G A G T G G  
C T G A G T C G A T T C G A G A C G G C G C T G C A C G C G C T G C G G A T C G C A A C C G T C G T C A T T C A C T G  
C T T C C G C T G C T G C A G A A C T A T C G A G A A C C C G C G A G G C G A T C C G G G G C G G C A T C G C G C C C  
G C A C C A C G G T T T C G C G G T G C G G T A C G G C A G G C G A A A A T C G G C C G G A C A A C G A C A T T C C C  
C A T G T C G G G C C C G G C A T C G C C A A G T A C G C C A G C G A C C T G C A G C T T C T C G G C T G G C T  
T G A

**FIG. 8TT****Amino acid sequence (SEQ ID NO:70)**

>uniprot|B2HE95|B2HE95\_MYCMM Fatty-acid-CoA ligase FadD9\_1

MSITCVDTRAQRSARRIEQLYSTDQFAAARPSTAVGIAISKSGLGLPQI  
IQTVMDGPQRPALGQRATRVVTDPNTGRSSAQLLAEFETITYRELWNRT  
NALTNAFAAEALADRGQRVCVLGFASIDYATIDLALMLLGAVSVPLPTNA  
ARAQLCHIVSETQPSLIASSTENLPDAISLVLSHRAPHRVVFDYRPELD  
AHREALEAARARLAAIPVTVETLTAAIARGRTVRPAEADCGAQSAQADAPAL  
LIYTSGSTGPKGVVYTRNRVADFWRSTSKEVEATEQRTAPSITLFNMPM  
SHANGRQVLYGTLNSNGGTAYFTARSDLSTLFDLALVRPTELGFPPRIWD  
MLLERFGREVDRLRDGTAEGADPGALKARVAADLRQVLLGGRYALAMMG  
SAPISEQMKASVESLDDLVMEGYGSTEAGTVIINNEVQRQPVIDYKLVD  
VAELGYFLTDPRPYPRGEELLVKTRTLFSGYYRDPEDGAQVFDPDFYRTGD  
IMAQVGPDRLAYLDRRNVLKLSQGEFVAVSRLEAIFFANSPLVRQIFVYA  
NGARAYPLAVVVPQTDAQSRHGRAELKAEHTSLHRVAMSAGLAPYEIPR  
DFIVETTPFTPQNGLLTAIHKLARPHLTQRYGARLELLYTELADSQTRRL  
HRLRQTGGRLPALETIRRAAGALLGTETTEPRPEAHFKDLGGDSVSAVTF  
SNLLHDIFYGFDVPVGVLGPATDLRALASHVESRRGAGWGPSFASVHVP  
RATSVHAGDLKLAFLDTKTLAAATSLPAADARARTVLLTGATGFLGRYL  
VLEWLRRRLRAVGGKLICLVRRAASDEQARVRLDTAFDSGDPQLPEHFRQLA  
VDRLEVLAGDKSEPGGLDGPTWQRLADTVLIVDPATLVNHVLSYRQLF  
APNVAGTAELLRLALTTRKRPYAYVSTVSVANQIEPSAFTEDADIREISR  
TRTIDDSFANGYTTSKWASEVILREAHDLCLGPVTVFRCDMILADTSYAG  
QLNLADTFTRLMLSVAATGIAPASFYRLGPDGKRQPAHFDSLPEFIAEA  
VATLGARRHDFQVHHVANPHHDGVGLDEYVDWLVDAGCPIRRIPDYDEW  
LSRFETALHALPDRKRRHSLLLPLLQNYREPAEPIRGGIAPAPRFRGAVRQ  
AKIGRDNDIPHVGPAIIAKYASDLQLLGLA

**YP 001850422.1****Nucleotide sequence (SEQ ID NO:71)**

>uniprot|B2HN69|B2HN69\_MYCMM Fatty-acid-CoA ligase FadD9

ATGTCGCCAATCACCGCTGAAGAGCGGCTCGAGCGCCGCATCCAGGACCTTACGCCAAC  
GACCCCGCAGTTGCCGCCAAACCCGCCACGGCATCACCGCAGCAATCGAGCGGGCG  
GGTCTACCGCTACCCAGATCATCGAGACCGTCATGACCGGATACGCCATCGGCCGGCT  
CTCGCTCAGCGCTCGGTGAATTCTGACCGACGCCGGCACCCGCCACACCAGCTGCGA  
CTGCTCCCCACTTCGAACCATCAGCTACGGCGAGCTTGGGACCGCATCAGCGCACTG  
GCCGACGTGCTCAGCACCGAACAGACGGTGAACCGGGCGACCGGGCTGCTTGGTGGC  
TTCAACAGCGTCGACTACGCCACGATCGACATGACTTTGGCGCGCTGGCGCGGTGGC  
GTACCACTGCAGACCAGCGCGATAACCCAGCTGCAGCGATCGCTGCGCAGACCCAG  
CCCACCATGATCGCGGCCAGCGTCGACGCACTCGTACGCCACCGAATTGGCTCTGTCC  
GGTCAGACCGCTACCCGAGTCCTGGTGTGACCAACCGGCAGGGTGAACGCACACCGC  
GCAGCGGTGAAATCCGCCGGAGCGCCTGGCGGCTCGGCCGGTCGAAACCGTGGC  
GAGGCCATCGCGCGGGCGACGTGCCCCCGGGTGCCTCCGCCGGCTCGGCCGGCACC  
GATGTGTCGACGACTCGCTCGCCTACTGATCTACACCTCGGGCAGCACGGGTGCGCC  
AAGGGCGCGATGTACCCCGACGCAACGTTGCGACCTTCTGGCGCAAGCGCACCTGGTTC  
GAAGGGCGCTACGAGCCGTCGATCACGCTGAACCTCATGCCATGCAAGGCCACGTG  
CGCAAATCCTGTACGGCACGCTGTGCAATGGCGCACCGCCTACTTCGTGGCGAAAAGC  
GATCTCTCCACCTTGTGAAGACCTGGCGCTGGTGCAGGCCACCGAGCTGACCTCGTG

**FIG. 8UU**

CCGCGCGTGTGGACATGGTTCGACGAGTTCAGAGTGAGGTCACCGCCGCTGGTC  
GACGGCGCCGACCGGTGCCTCGAAGCCCAGGTCAAGGCCAGATAACGCAACGACGTG  
CTCGGTGGACGGTATACCAAGCGCACTGACCGGCTCCGCCCTATCTCGACGAGATGAAG  
GCGTGGGTCGAGGAGCTGCTCGACATGCATCTGGTCGAGGGCTACGGCTCCACCGAGGCC  
GGGATGATCCTGATCGACGGAGCCATTGGCGCCGGCGTACTCGACTACAAGCTGGTC  
GATGTTCCCGACCTGGGTACTTCCTGACCGACCGGCCACATCCGGGGCGAGTTGCTG  
GTCAAGACCGATAGTTGTTCCCGGGCTACTACCAAGCGAGCCGAAGTCACCGCCGACGTG  
TTCGATGCTGACGGCTCTACCGGACCGGCACATCATGGCCGAGGTGGCCCCAACAG  
TTCGTGTACCTGACCCGCCAACAACGTGTTGAAGCTGTCGAGGGCGAGTTGTCACC  
GTCTCCAAACTCGAACGGTGTGGCGACAGCCCCTGGTACGGCAGATCTACATCTAC  
GGCAACAGCGCCCGTGCCTACCTGTTGGCGGTGATCGTCCCCACCCAGGAGGCGCTGGAC  
GCCGTGCCTGTCGAGGAGCTCAAGGCAGGCTGGGCACTCGTCAAGAGGTCGCAAAG  
GCCGCCGCCCTGCAGTCTACCGAGATCCCGCGCACCTCATCGAAACAACACCATGG  
ACGCTGGAGAACGGCTGCTCACCGGACCGCAAGTTGGCCAGGCCAGCTGAAAAAG  
CATTACGGCGAGCTCTCGAGCAGATCTACACGGACCTGGCACACGCCAGGCCAGGAA  
CTGCGCTCGCTGCGCAAAGCGGTGCCGATGCGCCGGTGTGGTACGGTGTGCCGTGCG  
GCCGCCGCCCTGTTGGCGGCAGCGCTCTGACGTCAGGCCAGCTGCGCACCTCACCGAT  
TTGGGCGCGACTCGCTGTCGGCGCTGTCGTTACCAACCTGTCGACGAGATCTCGAC  
ATCGAAGTGGCGGTGGCGTCACTCGAGCAGATCTACACGGACCTGGCACACGCCAGGCC  
TACGTCGAGGCCGCGCTCGCAAACCCGGCTCGTACGCCGACCTCGCCTCGGTCCACGGC  
GCCTCGAATGGCAGGTACCGAGGTGCACTGCCGCTGGGACAAATTCACT  
GATGCCGAAACCTGGCGAACAGCTCCCCGGCTGCCGCGAACACACCCAGTGCACCC  
GTGCTGCTGACCGGGGCCACCGGCTCCCTCGGGCGTACCTGGCCCTGGGATGGCTGGAG  
CGGATGGACCTGGTCGACGGCAAACGATCTGTCGCTGGTCCGGCCAAGTCCGACACCGAA  
GCACGGGCGGGCTGGACAAGACGTTGACAGCGGCCACCCGAACCTGGCCACTAC  
CGCGCAGTGGCGGCGACCACCTCGAGGTGCTGCCGGTACAAGGGCGAACGCCACCTC  
GGACTGGACCGGAGACCTGGCAACGCCCTGGCGACACGGTCGACCTGATCGTCAACCCCC  
GCCGCCCTGGTCAACCAACGACTGCCATACAGCCAGCTGTTGGGCCAACGCCGCTGGG  
ACCGCCGAGCTGTCGGCTGGCGCTCACCTCCAAGATCAAGCCCTACAGCTACACCTCG  
ACAATCGGTGTCGCCGACCAGATCCGCCGTCGGCGTTACCGAGGACGCCGACATCCG  
GTCATCAGGCCACCCGCCGCGTCGACGACAGCTACGCCAATGGCTACTCGAACAGCAAG  
TGGGCCGCGAGGTGCTGTCGCGAGGCCATGACCTGTCGCTGCCGGTTGCCGTGCG  
TTCGGCTGCGACATGATCCTGCCGACACCACATGGCGGGACAGCTCAATGTCGCCGAC  
ATGTTACCCGGATGATCCTGAGCCTGGCGGCCACCGGTATCGCGCCGGGTTGTTCTAT  
GAGCTTGGCGGCCACGGCGCCGGCAACCGGCCACTATGACGGTCTGCCGTCGAGTTC  
ATCGCCGAGGCGATTGACTTGGGTGCGCAGAGCCAGGATGGTTCCACACGTATCAC  
GTGATGAACCCCTACGACGACGGCATGGACTCGACGAGTTGCTGACTGGCTCAACGAG  
TCCGGTTGCCCCATCCAGCGCATCGCTGACTATGGCGACTGGCTGAGCGCTTCGAAACC  
GCACTGCCGCACTGCCGATGCCGAGCGGCCACAGCTCACTGTCGCGCTGTTGCAACAC  
TATCGGCAGCCGGAGCGGCCCTCCGCGGGTCACTGCCCTACCGATCGCTTCCGGGCA  
GCCGTGCAAGAGGCCAAGATGCCCTGACAAAGACATTCCGACGTCGGCGGCCGATC  
ATCGTGAAGTACGTCAGCGACCTGCGCCTACTCGGCCCTGCTCTGA

**Amino acid sequence (SEQ ID NO:72)**

>uniprot|B2HN69|B2HN69\_MYCMM Fatty-acid-CoA ligase FadD9

MSPITREERLERRIQDLYANDPQFAAKPATAITAALIERPGLPLPQIET  
VMTGYADRPALAQRSVEFVTDAIGTGHLLPHFETISYGEWDRISAL  
ADVLSTEQTVKPGDRVCLLGFNVDYATIDMTLARLGAVAVPLQTSAAIT  
QLQPPIVAETQPTMIAASVDALADATELALSGQTATRVLVFDHHRQVDAHR  
AAVESARERLAGSAVVELAEAIARGDVPRGASAGSAPGTDVSDDSLALL  
IYTSGSTGAPKGAMYPRRNVATFWRKRTWFEGGYEPSITLNFMPPSHVMG  
RQILYGTLCNGGTAYFVAKSDLSTLFEDLALVRPTELTFFVPRVWDMVFDE

FIG. 8VV

FQESEVDRRLVDGADRVALEAQVKAEIFRNDVLGGRTSALTGSAPISDEMKA  
AWVEELLDMHLDVYEGYSTEAGMILIDGAIIRPAVLVDVPLGYFLT  
DRPHPRGELLVKTDLSLFPGYQRAEVTA  
DVFADGFYRTGDIMAEGVGPEQF  
FYVYLDRRNNVLKLSQLGEFVTVSKLEAVFGDSPTI  
VRQTYIYGN  
SARAYLLA  
VIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQS  
YEIPRDFIIETTPW  
TLENGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSLRQSGAD  
APVLVTVCR  
AAAALLGGSASDVQPD  
AHFTD  
LGGDSLSALSFTNLLHEIFD  
IEPVGVIVSP  
ANDLQALADY  
VEAARKPGSSRPT  
FA  
SVHGASNGQVTEVH  
AGDLSLDKF  
IDAATLAEAPRL  
PAANT  
QV  
RTV  
LLTGATGFLGRY  
LA  
EWLE  
RMDLVDGK  
LICL  
VRAKSD  
TE  
RARLDKT  
FD  
SGDP  
ELLA  
HY  
RALAGDH  
LEV  
LAGDKGE  
ADI  
GLD  
RQT  
WQR  
LAD  
TV  
LIV  
DPA  
ALVN  
HVLP  
SQL  
FGPN  
ALG  
TAELL  
RLAL  
TSK  
IKP  
YS  
YT  
ST  
TIG  
VAD  
QI  
PPSA  
FTED  
AD  
IRV  
I  
SAT  
RAV  
DD  
SYANG  
YSNS  
KWA  
GE  
VLL  
RE  
A  
H  
D  
L  
C  
G  
L  
P  
VA  
F  
R  
C  
D  
M  
I  
L  
A  
D  
TT  
W  
A  
G  
Q  
L  
N  
V  
P  
D  
M  
F  
T  
R  
M  
I  
L  
S  
L  
A  
A  
T  
G  
I  
A  
P  
G  
S  
F  
Y  
E  
L  
A  
A  
D  
G  
A  
R  
Q  
R  
A  
H  
Y  
D  
G  
L  
P  
V  
E  
F  
I  
A  
E  
A  
I  
S  
T  
L  
G  
A  
Q  
S  
Q  
D  
G  
F  
H  
T  
Y  
H  
V  
M  
N  
P  
Y  
D  
D  
G  
I  
G  
L  
D  
E  
F  
V  
D  
W  
L  
N  
E  
S  
G  
C  
P  
I  
Q  
R  
I  
A  
D  
Y  
G  
D  
W  
L  
Q  
R  
F  
E  
T  
A  
L  
R  
A  
L  
P  
D  
R  
Q  
R  
H  
S  
S  
L  
P  
L  
L  
H  
N  
Y  
R  
Q  
P  
E  
R  
P  
V  
R  
G  
S  
I  
A  
P  
T  
D  
R  
F  
R  
A  
A  
V  
Q  
E  
A  
K  
I  
G  
P  
D  
K  
D  
I  
P  
H  
V  
G  
A  
P  
I  
I  
V  
K  
Y  
V  
S  
D  
L  
R  
L  
G  
L  
L

069484

### Nucleotide sequence (SEQ ID NO:73)

>uniprot|O69484|O69484 MYCLE Putative Acyl-CoA synthetase

ATGTCGACTATCACTAAGCAGAAAAGCAGCTCGCACGCCGCGTTGACGACCTCACCGCC  
AACGACCGCAGTCGCCGCCAAACCGACCCGGCGGTAGCCGCCGCCCTGCCAG  
CCCGGGCTTCGACTGCCCAAATCATCCAGACCGCGCTGGACGGTACCGGGAGCGGGCC  
GCACTGGGCCAGCGCTGCCAGTTACCAAAAGACCTAAGACCGGACGCACCTCGATG  
GAGCTGCTCCCCAGCTTGAGACCATCACCTACCGCCAGTTGGCGACCGTGTGGAGCG  
CTGGCGCGCCTGGAGGCACGACCTACTGCACGCCGCTACCGGGTCTGCGTGTAGGT  
TTCAACAGTGTGATTACGCCATCATCGACATGGCGCTCGCGTGATTGGTGCTGTGGCG  
GTTCCACTGCAGACCACTGCGCGATACCCAGCTGCGATCGTGTGACCGAGACCGAA  
CCCAGTATGATCGCGACGAGCGTAAACCAGCTGCCGATACTGTCGAGCTGATCCTGTCT  
GGCCAGGCGCCAGCGAAGCTCGTTGTGACTACCACCCCGAGGTGACGAGCAGCAT  
GACGAGTGGCAACCGCCGGCGCGTTGGCGACAGTAGCGTGTGGTGAGAGCCTG  
ACCGAGGTCTCGGCGGCAAGACGCTGCCAGCTACGCCGATCCCGTGGCGATGAC  
TCTGTCACCGTGGCGTTGCTGATCTACACATCTGGCAGCACCGCGCACCCAAAGGGC  
GCGATGTATCTGCAAAGCAATGTCGGCAAGATGTGGCGCGGTAGACGGAAACTGGTC  
GGGCCAACCGCCGCGTCAATCACTCTTAACTTCATGCCGATGAGCCACGTATGGCCGCG  
GGAATCCTCTACGGCACGCTCGTAACGGCGCACGGTTACTCGCCGCCGAGCGAC  
CTCTCGACGCTGCTGGAGGATCTCAAGCTGGTGCGCCGACCGAGTTGAACTTGTACCG  
CGCATCTGGAAACCTCTACGATGAATCCAACGCGCAGTTGACCGTGGTTAGCCAAC  
AGCGGCTCCGCCGACCGTGCAGCCATCAAAGCGAAGTTATGGATGAAACAGGCCAACCC  
CTGCTGGAGGACGGTACATCGCGCTATGACGGCTGGCGCAACCTCCCGGAGTTG  
AAACACGGGGTCGAGTCCCTACTCGAAATGCATCTGTTGAAAGGCTACGGCTCCACCGAA  
GCCGGCATGGTCTTGTGACGGCGAAGTGCAACGTCGCCGGTTATCGATTACAAGCTG  
GTCGACGTTCCGGATTGGCTACTCAGCACCGACCGCTTATCGAGAGGTGAATTG  
CTGCTCAAGACCCAGAACATGTTCCCCGGCTACTACAAGCGTCTGAGGTACCGCCACC  
GTGTTGACAGCGACGGTTACTACCAAGACCGGAGACATTGTCGCCGAGTCGGTCCCGAC  
CGGCTCGTGTACGTCGATGCCGCAACACGTGCTGAAACTCGCGCAGGGCCAGTTGCTC  
ACCGTCGCAACACTCGAGGCAGCGTTAGCAATAGCCACTGGTCCGGCAGATCTACATC  
TATGGCAACAGCGCACACCCCTACCTGTTGGCTGTTGGTGCGACCGAGGGATGCGTTG  
GCTACCAATGACATTGAGGTGCTAAACCGCTGATTATCGATTCTTACAGAAAAGTAGCG

**FIG. 8WW**

AAAGAAGCCGACCTGCAGTCCTACGAGGTGCCGCGACTTAATCGTCGAGACTACACCG  
 TTCAGCCTGGAGAAATGGCCTGCTCACCGTATTGCAAGCTGGCGTGGCGAAGCTCAAG  
 CAGCACTACGGCGCGACTCGAACAGCTCACGCCGATCTGGTTGAAGGTCAGGCAAAT  
 GCACTGCACGTGCTAAAACAAAGCGTGGCGACCGCTCCGGTACTGCAGACGGTGAGCCGA  
 GCCGTGGGCACCATTCTGGGAGTGGCGACCACCGATTGCCGTCGAATGCCACTTCACC  
 GACTTAGGAGGAGACTCGTTGCCGCGTGCACATTGGCTACGCCACTTCACGCCACTTTC  
 GACATCGATGTGCCGTGGCGTATTGTCAGCCGTCAACAACATTGGTGGCGATGCC  
 GACTACATCGAGCGCGAGCGCAGGGCACGAAGCGGCCACTTCATTGCCATACACGGT  
 CGTGACGCTGGCAAAGTGCATGCCAGTGACCTCACTCTAGACAAATTGATGTATCA  
 ACGCTGACTGCCGCCCGTATTGGCGAACCCGGACCGAGGTGCGCACCCTGCTGTTG  
 ACCGGCGTACCGGCTTCCTGGGCGCTACTTGGCCTGAAATGGCTCGAACGGATGGAC  
 CTGGTCGAAGGGAAGGTAATCGCTCTGGTAAGAGCCAAGTCCAACGAGGACGCTGGGCC  
 CGGCTCGACAAGACCTCGATAGCGGAGACCCAAACTGCTGGCGCACTACCAGGAACCTG  
 GCAACCGACCACCTGGAGGTCATGCCGGCGACAAAGCGAAGTAGATCTGGAATTGGAC  
 CGGCAAACGTGGCGACACTGCCGACACGGTCGATCTGATCGACCCGCCGCGCTG  
 GTCAACCACTGCTGCCGTACAGCGAGCTATTGCCCTAACGTTAGGCACCGCCAG  
 CTGATTGGATCGCCTGACCAGTAAGCAAAGCCGTACATCTATGTGTCGACAATGGC  
 GTCGGTAATCAGATTGAGCCACAAAATTCAACGAAAGACTCCGACATCCGAGTCATTAGC  
 CCGACGCCAACATCAACAACACTATGCCAACGGTACGGCACAGCAAGTGGCCGGC  
 GAAAGTGTGCTGCCGAAAGCTCACGACCTATGCCGCTGCCGTCACGGCTTCCGCTGC  
 GACATGATCTGGCCGACACCAGCTATGCCGCTAGCTAACGCTCCCCGACATGTTACT  
 CGAATGATGCTGAGTCTAGGCCACCAGGATCGCACCCGGCTCGTCTACGAGCTAGAC  
 GCCGAGAGCAATCGCAACGCCACTACGACGGTCTGCCGTCGAGTTCATGCCGAA  
 GCGATCTCACCCGGAGACCAAAGCCTGCACGATCGAGACGGTTCACGACCTATCAT  
 GTAATGAACCGCACGACGGCATCGGTATGGACGAGTTGACTGGTTATTGAT  
 GCCGGCTGCCCTATAACACGCAACGACTACGACGAATGGCTGCGACGGTTGAGATT  
 TCGCTGCCGCCCTGCCGAAGGGAGCGTCACAGCTACTGTTGCCGTTGTCACAC  
 TACCAAGCCGGAGAAGCATTGACCGGCTGGCACCCACAATCCGTTCCGTACG  
 GCCGTTCAAAACCGCAACATTGGTCAGGACAAAGATATTCCGATATCTGCCGGCAATC  
 ATGCCAAATATGTCAGCGATCTGAGCTGGCTGGTTGA

**Amino acid sequence (SEQ ID NO:74)**

>uniprot|O69484|O69484 MYCLE Putative Acyl-CoA synthetase

MSTITKQEQLARRVDDLTANDPQFAAKPDPAVAALAQPGLRLPQIIQ  
 TALDGYAERPALGQRVAEFTKDPKTGRSMELLPSFETITYRQLGDRVGA  
 LARAWRHDLHAGYRVCVLGFNSVDYAIIDMALGVIGAVAVPLQTSAAIT  
 QLQSIVTEPEPSMIATSVNQLPDTEVELILSGQAPAKLVVFDYHPEVDEQH  
 DAVATARARLADSSVVVESLTVLGRGKTLPATPIPVADDSDAPLALLIY  
 TSGSTGAPKGAMYLQSNVGKMWRRSDGNWFGPAAASITLNFMPPMSHVMGR  
 GILYGTGNGGTAYFAARSDLSTLLEDLKLRPTELFNPRIWETLYDES  
 KRAVDRRLANSGSADRAAIKAEVMDERQSQLLGGRYIAAMTGSAPTSPTEL  
 KHGVESLEMLLEGYGSTEAGMVLFDGEVQRPPVIDYKLVDVPDFGYFS  
 TDQPYPRGELLKLTQNMFPGYYKRPEVTATVFDSDGYYQTGDIVAEVGPD  
 RLVYDRRNNVLKLAQGQFVTVAKEAAFNSPLVRQIYIYGNSAHPYLL  
 AVVVPTEDALATNDIEVLKPLIIDSLQKVAKEADLQSYEVPRDLIVETTP  
 FSLENGLLTGIRKLAWPKLQHYGARLEQYADLVEGQANALHVLKQSV  
 NAPVLQTVSRAVGTILGVATTDLPNAHFTDIGGDSLSALTFGSLLRELF  
 DIDVPVGIVSPVNVLVAIADYIERERQGTRPTFIAIHGRDAGKVHASD  
 LTLDKFIDVSTLTAAPVLAQPGTEVRTVLLTGATGFLGRYLALKWLERMD  
 LVEGKVIALLVRAKSSEDARARLDKTFDSDGDPKLLAHYQELATDHLEVIAG  
 DKGEVDLELDRQTWRRLADTVDLIVDPAALVNHLVLPYSELFGPNTLGTAE  
 LIRIALTSKQKPYIYVSTIGVGNQIEPAKFTEDSDIRVISPTRNINNNYA

**FIG. 8XX**

NGYGN SKWAGEVLLREAHDL CGI.PVT VFR CDM ILADTS YAGQLN VPDM FT  
RMMLSLAATGIAPGSFYELDAESNRQRAHYDGLPVEFIAEAISTLGDQSL  
HDRDGFTTYHVMNP HDDGIGMDEFV DWLIDAGCPIQRINDYDEWLRRFEI  
SLRALPERQRHSSLLPLLHNYQKPEKPLHGSLAP TIRFRTAVQNANIGQD  
KDIPHISPAIIAKYVSDLQLLGLV

**Q10896**

**Nucleotide sequence (SEQ ID NO:75)**

>uniprot|Q10896|Q10896\_MYCTU PROBABLE PEPTIDE SYNTHETASE NRP (PEPTIDE SYNTHASE)

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAGCGACGCC  
CAGTCGCCGCCAGTCCCAACGAGGCATACCCAGGGCATCGACCAGCCCCGGTC  
GCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTACGCCGATGCCGGCAGTCCGC  
CAGCGTGCCTCGCTCGTACCGACCCGACAGCGGCCGACCATGGTCAGACTACTG  
CCGGGTTCGAGACCATCACCTACCGGA CTGTGGGCCGCGCCGACATTGCCACC  
GCGTTGAGCGCTGAGCCCGCATCCGGCCGGGAGCCGGGTTGGCGTGTGGCTTCAAC  
AGCGTCGACTACACAACC ATCGACATCGCCTGATCCGGTTGGCGCCGTGTCGGTTCCA  
CTGCAGACCAGTGC GCGCCGGTACCGGGTTGCGCCCGATCGTACCGAGACCGAGCCGACG  
ATGATGCCACCAGCATCGACAATCTGGCGACGCCGTCGAAGTGCTGGCCGGTACGCC  
CCGGCCCGGCTGGCTGTATTGATTACACGGCAAGGTTGACACCCACCGCGAGGCCGTC  
GAAGCCGCCGAGCTCGGTGGCGGCTCGGTGACCATCGACACACTGCCGA ACTGATC  
GAACCGGGCAGGGCGCTGCCGCCACACCCATTGCCGACAGCGCCGACGCGCTGGCG  
CTGCTGATTTACACCTCGGGTAGTACCGCGCACCCAAAGCGCCATGTATCGCGAGAGC  
CAGGTGATGAGCTTCTGGCGCAAGTCGAGTGGCTGGTCGAGCCGAGCGGTTACCCCTCG  
ATCACGCTGAACCTCATGCCGATGAGCCACGTGCGGGCCGTCAGGTGCTCTACGGGACG  
CTTCCAACGGCGGTACCGCCTACTTCGTCGCCAACAGAGCGACCTGTCACGCTGTTGAG  
GACCTCGCCCTGGTGC GGGCCACAGAATTGTCGTTCTCGTCCCGCGCATCTGGGACATGGT  
TTCGAGAGTCCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCGATCGAGCGCG  
CTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTCGCGGACGGTTGTGATG  
GCGCTGACCGGTTCCGCGCCGATCTCGCTGAGATGACGGCGTGGTCGAGTCCCTGCTG  
GCCGACGTGCATTGGTGGAGGGTTACGGCTCCACCGAGGCCGGGATGGTCTGAACGAC  
GGCATGGTGC GCGCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCAGCTGGC  
TACTCGGCACCGATCACGCCCTACCCCGGGCGAGCTGCTGGTCAAGACGCAAACCATG  
TTCCCGGCTACTACCAGCGCCGGATGTCACCGCCGAGGTGTTGACCCCGACGGCTTC  
TACCGGACGGGACATCATGCCCAAAGTAGGCCCGACCAGTTCGCTCACCTCGACCGC  
CGCAACAACGTCTAAAGCTCTCCAGGGCGAGTTCATGCCGTCGAAGCTCGAGGCG  
GTGTTCGCGACAGCCCGCTGGTCCGACAGATCTCATCTACGGCAACAGTGCCGGGCC  
TACCCGCTGGCGGTGGTTGTCGGGACGCCGCTTCTGCCATGGCATCGAGAAT  
CTCAAGCCGTGATCGCAGTCCCTGCAGGAGGTAGCGAGGGCGGCCCTGCAATCC  
TACGAGATTCCACCGCAGCTCATCGAAACCACGCCGTCACCTGGAGAACGGCTG  
CTCACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCAACGTCTC  
GAGCGGCTCTACCGAGCTGGCGATAGCAATCCAACGAGCTGCCGAGCTGCCGAA  
AGCGGTCCCGATGCCCGGTGCTTCCGACGCTGTGCCGTGCCGCGCTGCCGTTGCTGGC  
TCTACCGCTGCCGAGTGTGCCGCCAGCGCAGCTCGCCGACCTGGGTGGTACTCGCTC  
TCGGCGCTGCTGTTGGCCAACCTGCTGTCAGAGATCTCGCGTCGACGTGCCGGTGGT  
GTCATTGTCAGGCCGGCAAGCGACCTGCCGCCCTGGCCGACCACTCGAAGCAGCGC  
ACCGCGTCAGGCAGCCAGCTCGCCTCGATAACCGGTGCTCCGCCAGGAAAGTGCAC  
GCCAGCGACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCCTGCCGAGCCCCAAC  
CTGCCGGCACCGAGCGCCCAAGTGC CGCACCGTACTGCTGACCGGCCACCGGCTTTG  
GGTCGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTGATC

**FIG. 8YY**

TGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCCGCTGGACGCGACGTTGAT  
AGCGCGACCCGTATTGGTGC GGCACTACCGCGAATTGGCGCCGGCGCCTCGAGGTG  
CTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCGCTGGACCGGGTCACCTGGCAGCGGCTA  
GCCGACACGGTGGACCTGATCGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTAT  
AGCCAGCTGTTGGCCCAAACCGCGCCGGCACCGCCGAGTGCTTCGGCTGGCGCTGACC  
GGCAAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTTGGCGAGCAGATCCCACCG  
GAGGCCTTACCGAGGACGCCGACATCGGGCCATCAGCCGACCCGAGGATCGACGAC  
AGCTACGCCAACGGCTACCGGAACAGCAAGTGGCCGGCAGGTGCTGCTGCCGAAGCT  
CACGAGCAGTGGCCCTGCCGTGACGGTCTCGCTGCGACATGATCCTGGCCGACACC  
AGCTATAACCGGTCAACCTGCCGGACATGTTCACCCGGCTGATGCTGAGCCTGGCC  
GCTACCGGCATCGCACCCGGTCTGTTCTATGAGCTGGATGCGCACGGCAATCGGAACGC  
GCCCACTATGACGGCTTGGCGTCAATTGCTGCGAGAACCAATTGACCCCTGGGACA  
CATAGCCCGGACCCTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG  
CTGGACGAGTTCGTCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCACGATCCAG  
CGGATCGCCGACTACGGCAGTGGCTGCAGCGGTTGAGACTTCGCTGCGTGCCTGGCG  
GATGCCAGGCCACGCGCTGCTGCCCTGCTGCACAACCTACCGAGAGCCTGCAAAG  
CCGATATGCGGGTCAATCGCGCCACCGACCAGTCCGCGTGCCTCCAAGAACGAAA  
ATCGGTCCGGACAAAGACATTCCGCACCTACGGCGGCGATCGCGAAGTACATCAGC  
AACCTGCGACTGCTCGGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:76)**

>uniprot|Q10896|Q10896\_MYCTU PROBABLE PEPTIDE SYNTHETASE NRP (PEPTIDE SYNTHASE)

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM  
EGYADRPALGQRALRFVTPDSGRTMVELLPRFETITYRELWARAGTLAT  
ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSPLQTSAPVTGL  
RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVFDYHGKVDTREAV  
EAARARLAGSVTIDTLELIERGRALPATPIADSADDALALLIYTSGSTG  
APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPPMSHVGGRQVLYGT  
LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR  
RLVDGADRAALEAQVKAELENRLENVLGGRFVMAUTGSAPIAEMTAWVESLL  
ADVHLVEGYGSTEAGMVNLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR  
GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMAKVGPDQFVYLD  
RNNVLIKLSQGEFIAVSKLEAVFGDSPLVRQIFIFYGNSARAYPLAVVPSG  
DALSRHGIENLKPVISESILQEVARAAGLQSYEIPRDFIIETTPFTLENGL  
LTGIRKLRPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT  
LCRAAAALLGSTAADVRPAHFADLGGDSLSALSLANLLHEIFGVDPVVG  
VIVSPASDLRALADHIEARTGVRRPSFASIHRGSADEVHASDLTLDKF  
DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI  
CLVRARSDEEAQARLDATFDSDGPYIIVRHYRELGAGRLEVLAGDKGEADL  
GLDRVTVQRLADTVLIVDPAALVNHLPPSQLFGPNAAGTAELLRLALT  
GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAIISPTRRIDDSYANGYANSK  
WAGEVILLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA  
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH  
VMNPYDDGIGLDEFVDWLNSPSGSGCTIQRIADYGEWLQRFETSLRALP  
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPH  
TAAIIAKYIISNLRLLLG

## FIG. 8ZZ

**Q5YY80****Nucleotide sequence (SEQ ID NO:77)**

&gt;uniprot|Q5YY80|Q5YY80\_NOCFA Putative carboxylic acid reductase

GTGGAATCCACACGAGCGACGCGTCTCGGGCAGCGGATCGCCGCCCTGTACGCCGACGAC  
GCGCAGGTGCGCGACGCGCGGCCGACGAGGCGATCAGCACGGCCTGCAGGGAAACCAGGGC  
CTGCGGCTGCGCGAGCTCGTCGCCACCGTGGCTGACGGCTACCGCGACCGGGCGCTG  
GCCGCGCCTCGGTGAGCCGGCGTCAGCAGCGACCGGGTGCCTGCGTTGCGCGCGTGG  
CTGCCCAGAGTACACGACGATGAGTTACGGCAGCTCGCCTGCGTTGCGCGCGTGG  
GCGGCCTGGCAGCACGACGAAACCCGCTTCGCCCGGGCAATTGTCGCGACCCCTG  
GGTTTACCAAGCCCCGACTACGCCGTGTCGACCTGGCTGCGTGTGGCGGGCGCGTGG  
GCGGTGCCGCTGCAGGGCGAGCGCGTCGGTGACGCACTGACCGCCATCCTGCCGAGACC  
GOGCCCGCAATCTGGCACCGGCCACCGGCTGGACACGCTGCCGACGCCGTTGACTGTGCTC  
GCCGGGGCACGCCACGGCACTGCACGTGTTGACTTCGACCCGCCATCGACGCGCAG  
CGCACGGTGTACGAGGCCGTGTCGCGGGTGGCGTACCGGTGTGCGCGTGCACG  
CTCGCCGAGGTCGAGGACCGCGGCCGGCGTGCCTGCGTATCGACGACGCC  
GGCGACGACCGCTGCCCTGTTGATCTACACCTCCGGCAGTACCGCACGCCAAGGGG  
GCGATGTACACCGAGCGGCTGGTCGCGCTGATGTTGCTGGCGAGCCGAGGTCGCCGCG  
CTCACCGTCAACTACCTGCCGCTCAGGCCACGTCGCCGGCGCTGGCGTGTTCGGCGT  
CTCGCGCGGGCGGACCGCCTACTTCACCGCGCGGCCACATGTCACGCTGTTGAG  
GATCTGGCGTGGCCAGGCCGACCGAGCTGTTGCTGGTGCCGCGTGTGCGAGATGGT  
CTGCAACGATTCCAGACCGAGCGCTGCGGCCAGGGCAGCACGACCGGGTCAAGGCC  
GACCTGCGCCTCGAACCTGTCGACCGGGCTGCTCTCGTGGTGCGGCCAGCGCG  
CTGGCCCCGGAGCTGAAGGCGTTATGGAATCGGTGCTGACCTGACGACGCC  
TACGGCTCCACCGAGGCCGGCGCAGCGTGGTCATCGACACCACCGTGCAGGCC  
GTGCTGGACTACCGGCTCGCCGACGTGCCCCACTGGCTATTCCGTACCGACAAGGCC  
CATCCGCGCGGCCGAGCTGCTGCTCAAGACCACCATGATCCCCGGCTACTACCGCG  
CCGGAGCTAACGCCAGATCTCGACGAGGACGGCTTCTACCGCACCGCGACGTGGTC  
GCGGAACCTGGCGCCGACCGGCTCGTGTACGTCATGCCGCAACAATGTGCTCAAGCTG  
GCGCAGGGCGAGTTGTCACCATGCCCGGCTGGAGGCATCTCGCCAACAGTCCGCTG  
GTGCGCCAGATCTCGTCTACGGCAACAGCGAACCGCCTATCTGCTGGCGGTATCGT  
CCGAGCGACAGGCATGGCGGGGATCCGCCACGCTGAAGACCGGGATCGCGGAGTC  
TTGCAAGCTCATGGCGGGGACGCCGAGCTGGAGGCCTACGAGATCCCGCGACTTCC  
ATCGAGACCGAGCCGTTACCAACCGAATCCGGCTGCTCTCGGGCATCGGCAAGATC  
CGTCCCGCCGCTCGAGGCCGCTATCGCACCGGCTCGAACAGCTACGCCACCTGG  
GCCGCCAGAGCTGGCGGCTGCCCGAGGCCGGCAGCGTCCGGTGC  
GAGACCGTCACCGCGGCCGCGATCCCGTGGCGGCTGCCCGAGGCCGGCAGCG  
GCCGCGCAGCTACCGATCTCGGCCGATTGCTGGCGCTGGCGTGTGCAACCTG  
CTCGGTGAGATCTCGCGTCAAGGAGTGGCGGCCACCGGCCGCGCCGGTGC  
CTCCGTTGGCTGGCGCCACATCGCGCGAACCGAAAACCGCACCGAGACACCG  
TTCGACCGGGTGCATCCGACCAAGATCCTGATCCGGGCCACCGACCTGCC  
TTCTCGACGCCGAGGAGTTGGCGGCCACCGGCCGCGCCGGTGC  
CGGGTGGTGCCTGACCGGTGCCAACGGCTATCTCGGCCGGTTCTGTGC  
CTGGAACGGCTCGACCGCGTCACGGACGGCTGATCTGCC  
GCCGCCGCGCTGGCGGCCCTGGAAAGCCGCTTCGACAGCGCG  
CGCTCAAGGAATTGGCCAGCGCAGGCTCACCGTGGTGCGCG  
GCCCTGGCGCTGGCCACCGGCCACGTGGCGACGGCTGCC  
CACCGGCCGCGCTGGCAACCGAACCGTGGCTGCC  
GCGGGCACCGCGAGATCCTGCGCTCGCGTACCGAGCG  
CTGTCACGGTGCCTGCGCAGATAACCGGCCGACCG  
ATCCCGTGTACAGCCGACCCGACGGTGGACCGCG  
AGCAAATGGCCGCGCAGGTGTTGCTGCGTGC  
AGCACCGCTCGATCTCCGGT

**FIG. 8AAA**

GGCGTGTCCGCTGGACATGATCCTGGCCACGGCAGCTCGCCGGACAGCTAACATC  
 CCCGACGTGTTCACCGGCTGCTGCTCAGCCTGCTGGTACCGGTATCGCGCCCGCTCG  
 TTCCACGCCGCACGGTCACCGCGAGCGCCCGCCACTACGACGGGCTGCCCGCG  
 GACTTCACCCTGCCCGATCACCGCCTGGGGCGCACCACGGGATTCCACACCTAC  
 GACGTGCTCAACCGCAGCAGCAGGCATCAGCCTGGACACCTCGTGGACTGGCTGATC  
 GAGGCCGGACATCCCATCGAACGCATCCCCGAGCACAGCAGTCAGCGAAACACTCGTGC  
 ACGGCGTTGCACGCCCTGCCGAACGTCAAGCGAAACACTCGTGCCTCCGCTGTCAC  
 GCCTACCGCAGGCCGGTGCCCGCCTGCCGGCTGCCGCTGCCGCCGGAGTCCGG  
 GCCGGCGGTGCGGGCCGCAGGCATCACCGCCACGGTACATCCCCCACCTGACGCCGCG  
 CTGATCGAGAAGTACGTCGCCGATCTCCGCTGCACGGACTGTTGTAG

**Amino acid sequence (SEQ ID NO:78)**

>uniprot|Q5YY80|Q5YY80\_NOCFA Putative carboxylic acid reductase

VESTRATRLQRRIAALYADDAQVRDARPDEAISTALREPGLRLRELVATV  
 VDGYRDRPалаарsvqavdaatgacvarllpeyttmsygelglrlrava  
 AAQHDDETRLRPGEFVATLGFTSPDYAVVVLACWAGAVAVPLQASASV  
 TQLTAILAETAPAILATGLDTLPHAVDCVLAGATPRAHVFDFDAIDAQ  
 RTVYEACARLAGTGVVRVTLAEVEDRGRALEPPAVIDDPGDDRLALLIY  
 TSGSTGTPKGAMYTERLVALMWLGQPQVAALTVNYLPLSHVAGRILALFGL  
 LARGGTAYFTARADMSTLFEDLALARPTELFVVPRVCEMVLQRFQTERLR  
 RQADDDRVKAADRLELFGDRLLSVVCGSAPLAPELKAFMESVLDLTLHDG  
 YGSTEAGGSVVIDTWRPPVLDYRLADVPPELGYFRDKPHPRGELLKT  
 TTMI PGYYRPELNAQIFDEDGFYRTGDVVAELAPDRLVYVDRNNVLKL  
 AQGEFTIARLEAIFANSPLVRQIFVYGNSERAYLLAVIVPSRQAMAGDP  
 ATIKTRIAESLQLIGRDAELEYEIPRDFLIETEPFTTESGLLSGIGKIL  
 RPAVEARYRDRLEQLYADLAQQDELAALRREAGQRPVLETVTAAAI  
 LGGTASDLSPAHHFTDGGDSLAAALALSNLREIFAVEVPVGVIITGPATD  
 LRGLAAHTAAERENRRTPLFDRVHPDQILIRATDLALEKFFDAEELAAA  
 ATAAPPVAEPRVVLTTGANGYLGRFLCLEWLERLDRVDGRLICLVRGADE  
 AAALARLEAAFDSGDPVELVRRFKELAQRRLTVVAGDIGEPGLGLATATWR  
 RLAAEVHEIVHPAALVNHLVLPYRQLFGPNVAGTAEILRLALTERKPIDF  
 LSTVAVAQQIPADRFaedGDIRVISPTRTVDRGYANGYGN SKWAAEVLLR  
 AAHDRFDLPAVFRSDMILAHSFAGQNLIPDVFRLLLSSLVTGIAPAS  
 FHAATVTGERPRAHYDGLPADFTAAAITALGARTAGFHTYDVLNP HDDGI  
 SLDTFVDWLIEAGHPIERIPEHSEWVTRFETALHALPERQRKHSILLPLLH  
 AYRRPVPALRGSSALPAAEFRAAVRAAGITADGDIPHLLTRALIEKYVADLR  
 LHGLL

**Q6RKB1****Nucleotide sequence (SEQ ID NO:79)**

>uniprot|Q6RKB1|Q6RKB1\_9NOCA ATP/NADPH-dependent carboxylic acid  
 reductase

ATGGCAGTGGATTACCGGATGAGCGGCTACAGCGCCGCATTGCACAGTTGTTGCAGAA  
 GATGAGCAGGTCAAGGCCGACGTCCGCTCGAAGCGGTGAGCGCCGGTGAGCGCGCC  
 GGTATGCGGCTGGCGCAGATCGCCGCACTGTTATGGCGGGTACGCGACCGCCGG  
 GCCGGGCAGCGTGCCTCGAACTGAACACCGACGACGCCGACGGGCCACCTCGCTGC

**FIG. 8BBB**

TTACTTCCCCGATTGAGACCATCACCTATCGGAACGTGGCAGCGAGTCGGCGAGGTT  
GCCGC GG CCT TGG CAT CAT GAT CCC GAG A ACC C CT TGC CG CAG GT GAT T C G C C C T G  
CT CG G C T C ACC A C AG C AT CG ACT AC G CC ACC C T CG AC CT GG CG AT AT C C ACC T CG G C G  
GTT ACC CGT G C G GT T G CAG G C AG CG C G G GT G T C C AG CT G AT CG C T AT C C T C ACC G AG  
A C T T C G C C G G C T G C T C G C T C G A C C C C G G AG C AC C T CG AT G C G C G G T C G A G T G C T A  
C T C G C G G C ACC A C ACC G G A A C G A C T G G T G G T C T C G A C T A C C A C C C G G A G G A C G A C G A C  
C A G C G T G C G G C T T C G A A T C C G C C G C C G C C T G C C G A C G C G G C A G C T T G G T G A T C  
G T C G A A A C G C T C G A T G C C G T G C G T G C C C G G C G A C T T A C C G G C G C C A C T G T T C  
G T T C C C G A C A C C G A C G A C G A C C C G C T G C G C T G A T C A C A C C T C G G C A G C A C C G G A  
A C G C C G A A G G G C G C G A T G T A C A C C A A T C G G T T G G C C G C A C G A T G T G G C A G G G G A A C T C G  
A T G C T G C A G G G A A C T C G C A A C G G G T C G G G A T C A A T C T C A A C T A C A T G C C G A T G A G C C A C  
A T C G C C G G T C G C A T A T C G T G T C G G C G T C G C T C G C G G T G G C A C C G C A T A C T C G C G  
G C C A A G A G C G A C A T G T C G A C A C T G T C G A A G A C A T C G G C T T G G T A C G T C C C A C C G A G A T C  
T T C T C G T C C C G C G C G T G C G A C A T G G T C T T C C A G C G C T A T C A G A G C G A G C T G G A C C G G  
C G C T C G G T G G C G G G C C G A C C T G G A C A C G C T C G A T C G G G A A G T G A A A G C C G A C C T C G G  
C A G A A C T A C C T C G G T G G G C G C T C C T G G T G G C G G T C G C G C A C C G C C C G C T G G C C G C  
G A G A T G A A G A C G T T C A T G G A G T C C G T C T C G A T C T G C C A C T G C A C G A C G G G T A C G G G T C G  
A C C G A G G C G G G C G C A A G C G T G C T G C G A C A C C A C G A T C C A G C G G C C G C T G C G A T  
T A C A A G C T C G T C G A C G T G C C C G A C T G G G T T A C T C C G C A C C G A C C G C C G C A T C C G C G C  
G G T G A G G C T G T G T G A A G G C G G A G A C C A C G A T T C C G G G C T A C T A C A A G C G G C C C G A G G T C  
A C C G C G G A G A T C T C G A C G G A C G G G C T T C A C A A G A C C G G C G A T A T C G T G G C C G A G G C T C  
G A G C A C G A T C G G C T G G T C T A T G T C G A C C G T C G C A A C A A T G T G C T C A A A C T G T C G C A G G G C  
G A G T T C G T G A C C G T C G C C C A T C T C G A G G C C G T G T C G C C G A C G C A G C C C G T G A T C C G G C A G  
A T C T C A T C T A C G G C A G C A G C G A C G T T C C T A T C T G C T C G C G G T G A T C G T C C C C A C C G A C  
G A C G C G C T G C G C G G C G C G A C A C C G C C A C C T T G A A A T C G G C A C T G G C G A A T C G A T T C A G  
C G C A T C G C C A A G G A C G C G A A C C T G C A G C C C T A C G A G A T T C C G C G C G A T T T C C T G A T C G A G  
A C C G A G C C G T T C A C C A T C G C C A A C G G A C T G C T C T C C G G C A T C G C G A A G C T G C T G C G C C C  
A A T C T G A A G G A A C G C T A C G G C G C T C A G C T G G A G C A G A T G T A C A C C G A T C T C G C G A C A G G C  
C A G G C C G A T G A G C T G C T C G C C C T G C G C C G A A G C C G C C G A C C T G C C G G T G C T C G A A A C C  
G T C A G C C G G G C A C G C G A A A G C G A T G C T C G G C G T C G C C C T C C G C G A T A T G C T C C C G A C G C G  
C A C T T C A C C G A C C T T G G G C G G C A T T C C C T T C C G C G T C G C G T C T C G A A A C T G C T G C T G C A C  
G A G A T C T C G G G G T C G A G G T G C C G G T G G G T G C G C G T C G C G C G A A C G A G C G A C T G C G C  
G A T C T G G C G A A T T A C A T T G A G G C G G A A C G C A A C T C G G G C G C G A A G C G T C C C A C C T T C A C C  
T C G G T G C A C G G C G G C G G T T C C G A G A T C C G C G C C G C G A T C T G A C C C T C G A C A A G T T C A T C  
G A T G C C C G C A C C C T G G C G A C A G C A T T C C G C A C G C G C C G G T G C C A G C G C A G A C G C  
G T G C T G C T G A C C G G C G C G A A C G G C T A C C T C G G C C G G T T C C T G T G C C T G G A A T G G C T G G A G  
C G G C T G G A C A A G A C G G G T G G C A C G C G C T G A T C T G C G T C G C G C G T A G T G A C G C G G C C G C G  
G C C C G T A A A C G G C T G G A C T C G G C G T T C G A C A G C G G C G A T C C C G C C T G C T C G A G C A C T A C  
C A G C A A C T G G C G C A C G G A C C C T G G A A G T C C T C G C C G G T G A T A T C G G C G A C C C G A A T C T C  
G G T C T G G A C G A C G C G A C T T G G C A G C G G T T G G C G A A A C C G T C G A C C T G A T C G T C C A T C C C  
G C C G C G T T G G T C A A C C A C G T C C T T C C C T A C A C C C A G C T G T C G G C C C C A A T G T C G T C G G C  
A C C G C C G A A A T C G T C C G G T T G G C G A T C A C G G C G C G C A A G C C G G T C A C C T A C C T G T C G  
A C C G T C G G A G T G G C C G A C C A G G T C G A C C C C G G A G G T A T C A G G A G G A C A G C G A C G T C C G C  
G A G A T G A G G C G C G G T G C G C G T C G C G C G A A G G A G A G T T A C G C C A A C G G C T A C G G C A A C A G C A A G  
T G G G C G G G G G A G G T C C T G C T G C G C G A A G C A C A C G A T C T G T G G C T T G C C G G T C G C G G T G  
T T C C G T T C G G A C A T G A T C C T G G C G C A C A G C C G G T A C C G G G G T C A G C T C A A C G T C C A G G A C  
G T G T T C A C C C G G C T G A T C C T C A G C C T G G T C G C C C A C C C G G A T C C G C C G C T A C T C G T T C T A C  
C G A A C C G A C G C G G A C C G G A A C C C G G A C G C G G G C C C A C T A T G A C G G C T T G C C G G C G G A C T T C  
A C G G C G G C G G C G A T C A C C G C G C T C G G C A T C C A A G C C A C C G A A G G C T T C C G G A C C T A C G A C  
G T G C T C A A T C C G T A C C G A C G A T G G C A T C T C C C T C G A T G A A T T C G T C G A C T G G C T C G T C G A A  
T C C G G C C A C C C G A T C C A G C G C A T C A C C G A C T A C A G C G A C T G G G T T C C A C C G T T C G A G A G C G  
G C G A T C C G C G C G T G C C G G A A A A G C A A C G C C A G G C C T C G G T G C T G C C G T T G C T G G A C G C C  
T A C C G C A A C C C T G C C C G G C G G T C C G C G G C G C G A T A C T C C C G G C C A A G G A G T T C C A A G C G  
G C G G T G C A A A C G C C A A A A T C G G T C C G G A A C A G G A C A T C C C G C A T T T G T C C G C G C C A C T G

**FIG. 8CCC**

ATCGATAAGTACGTCAGCGATCTGGAACTGCTTCAGCTGCTCTGA

**Amino acid sequence (SEQ ID NO:80)**

&gt;uniprot|Q6RKB1|Q6RKB1\_9NOCA ATP/NADPH-dependent carboxylic acid reductase

MAVDSPDERLQRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAAT  
VMAGYADRPAAGQRAFELNTDDATGRTSLRLLPRFETITYRELWQRVGEV  
AAAWHHDPEPPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQASAA  
VSQQLIAILTEETSPRLLASTPEHLDAAVECLLAGTTPERLVVF DYHPEDDD  
QRAAFESARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPDTDDDPLA  
LLIYTSGSTGT PKGAMYTNRLAATMWQGNNSMLQGNSQRVGINLNYPMSH  
IAGRIISLFGVLARGGTAYFAAKSDMSTLFEDIGLVRPTEIFFVPRVCDMV  
FQRYQSELDRRSVAGADLTLDRREVKADLRQNYLGGRFLVAVVGSA PLAA  
EMKTFMESVLDPLHDGYGSTEAGASVLLDNQIQRPPVLDYKLVDVPELG  
YFRTDRPHPRGEPLLKAETTI PGYYKRPEVTAEIFDEDGFYKTGDIV AEL  
EHDRLLVYDRRNNVLKLSQGEFVTVAHLEAVFASSPLIRQIFIYGS SERS  
YLLAVIVPTDDALRGRTATLKSALAESIQRIAKDANLQPYEIPRDFLIE  
TEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDLATGQADELLA LRR  
EAADLPVLETVSRAAKAMLGVASADMRPDAHFTDLGGDSLSALSFSNLLH  
EIFGVEVPVGVVVSPANELRDLANYIEAERNSGAKRPTFTSVHGGGSEIR  
AADLTLDKFIDARTLAAADSIPHAPVPAOTVLLTGANGYLGRFLCLEWLE  
RLDKTGGTLICVVRGSAAAARKRLDSA FDSDGDPGLLEHYQQLAARTLEV  
LAGDIGDPNLGLDDATWQRLAETVDLIVHPAAI.VNHVI.PYTQI.FGPNVVG  
TAEIVRLAITARRKPVTYLS TVGADQVDPAEYQEDSDVREMSAVRVVRE  
SYANGYGN SKWAGEVLLREAHDL CGLPVAVFRSDMILAH SRYAGQLNVQD  
VFTRLILS LVATGIAPYSFYRTDADGNRQRAHYDGLPADFTAAITALGI  
QATEGFRTYDVLPYDDGISLDEFVDWLVESGHPIQRTIDYSDWFHRFET  
AIRALPEKQRQASVPLLLDAYRNPCPAVRGAILPAKEFQAAVQTAKIGPE  
QDIPHLSAPLIDKYVSDLELLQLL

**Q741P9****Nucleotide sequence (SEQ ID NO:81)**

&gt;uniprot|Q741P9|Q741P9\_MYCPA FadD9

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGA ACTCATGCCACC  
GACCCGCAATT CGCCGCCGCCAACCCGACCCGGCGATCACCGCCGCCCTCGAACAGCCC  
GGGCTGCGCTGCCGCAGATCATCCGCACCGTGCTCGACGGCTACGCCGACCGGCCGGCG  
CTGGGACAGCGCTGGTGGAGTTCTCGCACGGACGCCAAGACCGGGCGCACGT CGCAG  
CTGCTCCCCGCTTCGAGACCATCACGTACAGCGAAGTAGCGCAGCGT GTTCGGCGCTG  
GGCCGCGCCCTGTCCGACGACGCGGTGCACCCGGCAGCGGTGTGCGT GCTGGGCTTC  
AACAGCGTCACTACGCCACCATCGACATGGCGCTGGCGCCATCGCGCCGTCTCGGTG  
CCGCTGAGACCAGCGCGCAATCAGCTCGCTGCAGCGATCGTGGCGAGACCGAGCCC  
ACCTGATCGCGTCCACCGTGAAACAGCTGTCGACGCCGTGCAGCTGATCACCGGCC  
GAGCAGGCCCAACCCGGCTGGTGGTCACTACCAACCGCAGGTGACGACCGAGCG  
GAGGCCGTCCAGGACGCCGCCGGCTGTCCAGCACCGCGTGGCCGTCCAGACGCTG  
GCCGAGCTGCTGGAGCGCGCAAGGACCTGCCCGTCCGGAGCCGCCGACGAG  
GACTCGCTGCCCTGCTGATCTACACCTCCGGTCCACCGGCCGGCAAGGGCGCGATG  
TACCCACAGAGCAACGTCGGCAAGATGTGGCGCCGCCGGCAGCAAGAACTGGT CGGCGAG  
AGCGCCGCCGTCGATCACCTGAACCTCATGCCGATGAGCCACGTGATGGGCCGAAGC ATC

**FIG. 8DDD**

CTCTACGGCACGCTGGCAACGGCGCACCGCCTACTTCGCCGCCGCAGCGACCTGTCC  
 ACCCTGCTTGGAGCACCTCGAGCTGGTGCAGGCCAACCGAGCTAACCTCGTCCCAGGATC  
 TGGGAGACGCTGTACGGCAATTCCAGCGTCAGGTGAGCGGGCTCTCCGAGGCCGG  
 GACGCCGGCAACGTGCGCCCGTCAAGGGCAGGTGCTGGCCAGCAGGCCAGTACCTG  
 CTGGCGGGCGGTTACCTTCGCGATGACGGCTCGGCCCATCTGCCGGAGCTGCGC  
 AACTGGGTCGAGTCGCTCGAAATGACACCTGATGGACGGTACGGCTCCACCGAGGCC  
 GGAATGGTGTGTTGACGGGAGATTCAAGCGCCGCCGGTATCGACTACAAGCTGGTC  
 GACGTGCCGGACCTGGGCTACTCAGCACCGACCAGGCCATCCGCGGGAGCTGCTG  
 CTGCGCACCGAGAACATGTTCCGGGCTACTACAAGCGGGCAAACCCACCGCGGGCGTC  
 TTCGACGAGGACGGTACTACCGCACCGGGGACGTGTTGCGCCAGATGCCCGGACCGG  
 CTGGTCTACGTCGACCAGCGCAACAACGTGCTCAAGCTGGCGAGGGCGAATTGTCAGC  
 CTGGCCAAGCTGGAGGCGGTGTCGGCAACAGCCGCTGATCCGCAAGATCTACGTCAC  
 GGCAACAGCGCCAGCCCTACCTGCTGGCGTGTGGTGCACCGAGGAGGCCGCTGCC  
 TCGGGTGACCCCGAGACGCTCAAGCCAAGATGCCGACTCGCTGCAGCAGGTGCCAAG  
 GAGGCCGGCCTGCACTACGAGGTGCCGCGACTTCATCATCGAGACCAACCCGTT  
 AGCCTGGAAAACGGTCTGCTGACCGGGATCCGGAAGCTGGCGTGGCGAAACTGAAGCAG  
 CACTACGGGAACGGCTGGAGCAGATGTACGCCGACCTGGCCGCCGGACAGGCCAACGAG  
 CTGGCCGAGCTGCCGCAACGGTGCCCAAGCGCCGGTGTGAGACCGTGAGGCCGCC  
 GCAGGCCGCACTGCTGGGTTGCGCCCTCCGACCTGCCCCGACGCCACTTCACCGAT  
 CTGGCGAGACTCGTTGTCGGCGTTGACATTGGCAACCTGCTGCGAGATCTCGAC  
 GTCGACGTCGCGGTAGCGTGTACGTCAGCCGCCAACGACCTGGGCCATCGCGAGC  
 TACATCGAGGCCGAGCGGCAAGGCAGCAAGGCCGACGTTGCCCTCGGACCGCAGGAG  
 GACCGCACCGTGGTGCAGGCCGACCTGACGCTGGACAAGTCCCTGACGCCGAGACG  
 CTGGCCGCCGCGCCGACCTGCCAAGCCGCCACCGAGGTGCCACCGTGTGCTGAC  
 GGCGCCACCGGCTTCTGGCCGTAACCTGGCCCTGGAATGGCTGGAGCGGATGGACATG  
 GTGGACGCAAGGTATGCCCTGGCCGGCTCCGACGAGGAGGCCAGCGCCGG  
 CTGGACAAGACCTCGACAGCGGCCACCGAAACTGCTCGCGCACTACCAGCAGCTGCC  
 GCCGATCACCTGGAGGTATGCCGCCGACAAGGGCAGGCCAATCTGGGCTGGGCAA  
 GACGTTGGCAACGACTGCCGACACGGTCACGTGATCGTCAACCCGCCGCTGGTC  
 AACACGTTGCCGATCACGACAGCTACGCCAACGGCTATGGCAACAGCAAGTGGGCCGG  
 ATCCGGCTGGCGCTGACGTCCAAGCAGAACGGCTACACCTACGTTGCCACCATCGCG  
 GGCGACCGAGATCGAGCCGGCAAGTCGTCAGAACGCCGACATCCGGCAGATGAGGCC  
 ACCCGGGCGATCAACGACAGCTACGCCAACGGCTATGGCAACAGCAAGTGGGCCGG  
 GTGCTGCTGCGCAGGCCGACGACCTGTCGGGCTGCCGCGTGTGAC  
 ATGATCCTGGCCGACACCACGTATGCCGGCAGCTCAACCTGCCGGACATGTTACCCGG  
 CTGATGCTGAGCCTGGTGGCACCGGGATCGGCCGGCTCGTCTACGAGCTCGACGCC  
 GACGGCAACCGGAGCGGGGCACTACGACGCCGCTGCCGCGTCAAGTCCAGACCTACCACGTG  
 ATCTCGACGCTGGGTTGCAAGATCACCGACAGCGACACCAGCTCCAGACCTACCACGTG  
 ATGAACCCCTACGATGACGCCGTCGGTCTGGACGAGTACGTCGATTGGCTGGAGGCC  
 GGCTATTGATCGAGCGGATGCCGACTACTCCGAATGGCTGCCGGTTGAGACCTCG  
 CTGCGGGCCCTGCCGACCGGAGCGCCAGTACTCGCTGCTGCCGCTGTCACAAC  
 CGCACGCCGGAGAACCGATCAACGGGTCGATAGCTCCACCGACGTGTTCCGGGAGCG  
 GTGCAGGAGGCGAAAATCGGCCCCGACAAGACATTCCGACGCTGTCGCCGCCGGTCATC  
 GTCAAGTACATACCGACCTGCAAGCTGCTCGGGCTGCTCTGA

**Amino acid sequence (SEQ ID NO:82)**

>uniprot|Q741P9|Q741P9\_MYCPA FadD9

MSTATHDERLDRRVHELIATDPFAAAQPDPAITAALEQPGLRLPQIIRT  
 VLDGYADRPALGQRVVEFVTDAKTGRTSQQLPRFETITYSEVAQRVSAL  
 GRALSDDAVHPGDRVCLGFNSVDYATIDMALGAIGAVSVPLQTSAAISS  
 LQPIVAEPTLIASSVNQLSDAVQLITGAEQAPTRLVVFDYHPQVDDQR  
 EAVQDAAARLSSTGVAVQTLAELLERGKDLPAVAEPPADEDSLALLIYT

**FIG. 8EEE**

GSTGAPKGAMYPQSNVGKMWRRGSKNWFGESAASITLNFMPPSHVMGRSI  
LYGTLGNNGTAYFAARSDLSTILLEDLELVRPTELNFVPRIWETLYGEFQR  
QVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPISPELR  
NWVESLLEMHLMGDYGSTEAGMVLFDGEIQRPPVIDYKLVDVPLGYFST  
DRPHPRGELLRTENMFPGYYKRAETTAGVFDEDGYYRTGDFFAEIAPDR  
LVYVDRNNVLKLAQGEFVTI LAKEAVFGNSPLIRQIYVYGNSAQPYLLA  
VVVPTEEALASGPETLKPKIADSLQQVAKEAGLQS YEVPRDFIIETTPF  
SLENLGTTGIRKLAWPKLQHYGERLEQMYADLAAGQANELAELRRNGAQ  
APVLQTVSRAAGAMILGSAASDLSPDAHFTDLGGDSLSALTFGNLLREIFD  
VDVPVGIVVSPANDIAAIASYIEAERQGSKRPTFASVHGRDATVVRAADL  
TLDKFLDAETLAAAPNLPKATEVRTVLLTGATGFLGRYLALEWLERMDM  
VDGKVIALVRARSDEEARARLDKTFDSCDPKLLAHYQQLAADHLEVIAGD  
KGEANLGLGQDWQRLADTVDVIVDPAALVNHLVLPYSELFGPNALGTAEL  
IRLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINSDYAN  
GYGNSKWAGEVLLREAHDLCLGPVAVFRCDMILADTTYAGQLNLPDMFTR  
LMLSLVATGIAPGSFYELDADGNRQRAHYDGLPVEFTAAAISTLGSQITD  
SDTGFQTYHVMNPYDDGVGLDEYVDWLVDAGYSIERIADYSEWLRRFETS  
LRALPDRQRQYSLLPLLHNYRTPEKPINGSIAPTDVFRAAVQEAKIGPDK  
DIPHVSPPVIVKYITDLQLLGLL

**Q7D6X4****Nucleotide sequence (SEQ ID NO:83)**

>uniprot|Q7D6X4|Q7D6X4\_MYCTU Substrate--CoA ligase, putative

ATGTCGATCAACGATCAGCGACTGACACGCCCGTCGAGGACCTATA CGCCAGCGACGCC  
CAGTTCCGCCGCCAGTCCCAACGAGGGATCACCCAGGCATCGACCAGCCC GGTC  
GCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTACGCCATCGGCCGGCACTCGGC  
CAGCGTGCCTCGCTCGTACCGACCCGACAGCGGCCGCACCATGGTCGAGCTACTG  
CCGCGGTTGAGACCACACCTACCGCGAACGTGGGCCCGCGCACATTGGCCACC  
GCGTTGAGCGCTGAGCCCGCATCCGCCGGCGACCAGGGTTTGCGTGCTGGCTTCAAC  
AGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGCGCCGTGCGTTCCA  
CTGCAGACCAGTGCGCCGGTACCGGGTTGCGCCGATCGTCACCGAGACCGAGCCGACG  
ATGATGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGTGCTGGCCGGTCACGCC  
CCGGCCCGGCTGGCGTATTGATTACACGGCAAGGTTGACACCCACCGCGAGGCCGTC  
GAAGCGCCCGAGCTCGGTTGCCGGCTCGGTGACCATCGACACACTTGCGAAGTATC  
GAACCGCGAGGGCGCTGCCGCCACACCCATTGCCGACAGCGCCGACGACGCGCTGGCG  
CTGCTGATTTACACCTCGGGTAGTACCGCGCACCCAAAGGCGCCATGTATCGCGAGAGC  
CAGGTGATGAGCTCTGGCGCAAGTCGAGTGGCTGGTCGAGCGGGTACCCCTCG  
ATCACGCTGAACCTCATGCCGATGAGCCACGTGGGGCCGTAGGTGCTACGGGACG  
CTTCCAACGGCGGTACCGCCTACTCGTCGCCAAGAGCGACCTGTCGACGCTGTCGAG  
GACCTCGCCCTGGTGCGCCACAGAATTGTGCTTCGTGCCGCGATCTGGGACATGGTG  
TTCGCAGAGTTCCACAGCGAGGTGACCGCGCCTGGTGGACGGCGGATCGAGCGCG  
CTGGAAGCGCAGGTGAAGGCCGAGCTGCCGGAGAACGTGTCGCCGGACGGTTGTGATG  
GCGCTGACCGGTTCCGCCGATCTCGCTGAGATGACGGCGTGGTCGAGTCCCTGCTG  
GCCGACGTGCATTGGTGGAGGGTTACGCCACCGAGGCCGGATGGTCTGAACGAC  
GGCATGGTGCGCCGCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGC  
TACTCGGCACCGATCAGCCCTACCCCCGGGCGAGCTGTCGGTCAAGACGCAAACCATG  
TTCCCGGCTACTACCGCGCCGGATGTCACCGCCGAGGTGTTGCGACCCGACGGCTTC  
TACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCAAGTTCGTCTACCTCGACCGC  
CGCAACACGTGCTAAAGCTCTCCAGGGCGAGTTCATGCCGTGTCGAAGCTCGAGGCG  
GTGTTCGCGACAGCCGCTGGTCCGACAGATCTTACGGCAACAGTGCCTGGC

**FIG. 8FFF**

TACCCGCTGGCGGTGGTTGTCCCGTCCGGGGACGCGCTTCCTGCCATGGCATCGAGAAT  
 CTCAAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCCTGCAATCC  
 TACGAGATTCCACCGACTTCATCATCGAAACCACGCCGTTCACCTGGAGAACGGCTG  
 CTCACCGCATCCGAAGCTGGCACGCCAGTTGAAGAAGTCTATGGGAACGCTCTC  
 GAGCGGCTCTATAACCGAGCTGCCGATAGCCAATCCAACGAGCTGCGCAGCTGCCAA  
 AGCGGTCCCAGTGCAGCGGTCTCCGACGCTGTGCCGTGCCGGCTGCCGTGGC  
 TCTACCGCTGCCAGTGTGCCGGACGCCACTTCGCCGACCTGGGTGGTGAECTCGCTC  
 TCGGCCGTGCTGGCAACCTGCTGACGAGATCTTCGGCGTCGACGTGCCGGTGGG  
 GTCATTGTCAGCCGGCAAGCAGCTGCCCTGCCGACCACATCGAAGCAGCGC  
 ACCGGCGTCAGGCAGCCAGCTCGCCTCGATACACGGTCGCTCCGCAGCGAAC  
 GCCAGCGACCTCACCGCTGGACAAGTTCATCGACGCTGCCACCTGGCGCAGCCCCAAC  
 CTGCCGGCACCGAGCGCCAAGTGCACCGTACTGCTGACCAGGCCACCGCTTTG  
 GGTCGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTGATC  
 TGCGCTGGTCCCGCAGATCCGACGAGGAAGCACAAGCCGGCTGGACGCGACGTTCGAT  
 AGCGGCAGCCGTATTGGTGCAGGCACTACCGCGAATTGGCGCCGGCCCTCGAGGTG  
 CTCGCCGGCAGAAGGGCAGGCCGACCTGGGCTGGACCGGGTACCTGGCAGCGGCTA  
 GCCGACACGGTGGACCTGATCGTGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTAT  
 AGCCAGCTGTTCGGCAAACCGCGGGCACCGCCGAGTTGCTCGGCTGGCGCTGACC  
 GGCAAGCGCAAGCCATACTACACCTCGACGATCGCGTGGCGAGCAGATCCGCCG  
 GAGGCAGTCACCGAGGACGCCGACATCCGGGCATCAGCCCACCCGAGGATCGACGAC  
 AGCTACGCCAACGGTACCGAACAGCAAGTGGCCGGCAGGTGCTGCTGCCGAAGCT  
 CACGAGCAGTGCAGGCTGCCGGTACGGCTTCCGCTGCGACATGATCCTGGCGACACC  
 AGCTATACCGGTACGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTGAGCCTGCC  
 GCTACCGGCATCGCACCCGGTCGTTCTATGAGCTGATGCCACGGCAATCGGCAACGC  
 GCCCACTATGACGGCTTGCCGGTCAATTGCTCGCAGAAGCCATTGCAACCTGGGACA  
 CATAGCCGGACCCTTGTACACACGTGATGAACCCCTACGACGACGGCATCGGG  
 CTGGACGAGTTCGTCAGCTGGCTCAACTCCCCACTAGCGGGTCCGGTTGCACGATCCAG  
 CGGATCGCCGACTACGGCAGTGGCTGCAGCGTTCGAGACTTCGCTGCCTGCCG  
 GATCGCCAGCGCCACGCCCTCGTGCCTGCAACTACCGAGAGCCTGCAAAG  
 CCAGATATCGGGTCAATCGGCCAACGACAGTCCCGCCTGCCGTCCAAGAAGCGAAA  
 ATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGCGATCATCGCGAAGTACATCAGC  
 AACCTGCGACTGCTCGGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:84)**

>uniprot|Q7D6X4|Q7D6X4\_MYCTU Substrate--CoA ligase, putative  
 MSINDQRIRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM  
 EGYADRPALGQRALRFVTDPDSGRTMVELLPRFETITYRELWARAGTLAT  
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAVTGL  
 RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAV  
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLITYTSGSTG  
 APKGAMYRESQVMSFWRKSSGWFEPSGPSITLNFMPPMSHVGRQVLYGT  
 LSNGGTAYFVAKSDLSTLFEDIALVRPTELCFVPRIWDMVFAEFHSEVDR  
 RLVDGADRAALEAQVKAELENVLGGRFVMAUTGSAPISAEMTAWVESLL  
 ADVHLVEGYGSTEAGMVLDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR  
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMKVGPDQFVYLD  
 RNNVNLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSG  
 DALSRHGIENLKPVISES LQEVARAAGLQSYEIPRDFIIETTPFTLENL  
 LTGIRKLARPQLKKFYGERLERLYTELADQSNELRELQSGPDAPVLPT  
 LCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDPVVG  
 VIVSPASDLRALADHIEARTGVRRPSFAI HGRSATEVHASDLTLDFKI  
 DAATLAAAPNLPAPSAQVRTVLLTGTGFLGRYLALEWLDRMDLVNGKLI  
 CLVRARSDEEAQARLDATFDSDGPYLVRYRELGAGRLEVLADKGEADL

**FIG. 8GGG**

GLDRVTVQRLADTVDLIVDPAALVNHLVPLSQLFGPNAAGTAELLRLALT  
GKRKPYIYTSTIAVGEQIRPEAFTEDADIRAIISPTRRIDSYANGYANSK  
WAGEVLLREAEAHEQCGLPVTVFRCDMILADTSYTQQLNLPDMFTRLMLSLA  
ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGHSPDRFVTYH  
VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP  
DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL  
TAAIIAKYISNLRLLGLL

**Q7TY99****Nucleotide sequence (SEQ ID NO:85)**

>uniprot|Q7TY99|Q7TY99\_MYCBO PROBABLE FATTY-ACID-CoA LIGASE FADD9  
(FATTY-ACID-COA SYNTHETASE) (FATTY-ACID-COA SYNTHASE)

ATGTCGATCAACGATCAGCGACTGACACGCCCGTCGAGGACCTATACGCCAGCGACGCC  
CAGTTCCGCCGCCAGTCCAACGAGGCATCACCCAGGCATGCCACCAGCCGGGGTC  
GCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTACGCCGATGGCCGGCACTCGGC  
CAGCGTGGCTCCGTCGACCGACAGCGCCGACCATGGTCGAGCTACTG  
CCCGGGTTGAGACCATCACCTACCGCAACTGTGGGCCGCGCCGACATTGGCCACC  
GCGTTGAGCGCTGAGCCCGCATCCGGCCGGCGACCGGGTTGCGTGCTGGCTTCAAC  
AGCGTCGACTACACAACCATCGACATCGCCTGATCCGGTTGGCGCCGTGTCGGTTCCA  
CTGCAGACCACTGCGCCGGTACCGGGTTGCGCCCGATCGTACCGAGACCGAGCCGACG  
ATGATGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGTGCTGGCCGGTACGCC  
CCGGCCCGGGCTGGTCGATTGCGATTACACCGCAAGGTTGACACCCACCGCGAGGCCGTC  
GAAGCCGCCGAGCTGGTTGGCCGGCTCGGTGACCATCGACACACTTGCGAACTGATC  
GAACCGCCGAGGGCGCTGCCGCCACACCCATTGCCGACAGCGCCGACGCCGCTGGCG  
CTGCTGATTACACCTCGGGTAGTACCGCGCACCCAAAGGCATGTATCGCGAGAGC  
CAGGTGATGAGCTCTGGCGCAAGTCGAGTGGCTGGTCAGGCCGAGCGGTTACCCCTCG  
ATACGCTGAACCTCATGCCGATGAGCCACGTCGGGGCCGTCAAGTGTCTACGGGACG  
CTTCCAACGGCGGTACCGCCTACTACGTCGCCAAGAGCGACCTGTCGACGCTGTTGAG  
GACCTCGCCCTGGTGCGGCCACAGAATTGTGCTTGTGCGCGCATCTGGGACATGGTG  
TTCGAGAGTCCACAGCGAGGTGACCGCCGTTGGTGGACGGCCGATCGAGCGCG  
CTGGAAGCGCAGGTGAAGGCCGAGCTGCCGGAGAACGTGCTCGGGGACGGTTGTCATG  
GCGCTGACCGGTTCCGCGCCGATCTCCGCTGAGATGACGGCGTGGTOGAGTCCCTGCTG  
GCCGACGTGATTGGTGGAGGGTATCGGCTCCACCGAGGCCGGGATGGTCTGAACGAC  
GGCATGGTGCAGCGCCCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGG  
TACTCGGCACCGATCAGCCCTACCCCGGGCGAGCTGCTGGTCAAGACGCAAACCATG  
TTCCCGGCTACTACCAAGCGCCGGATGTCACCGCCGAGGTGTTGCGACCCGACGGCTTC  
TACCGGACGGGACATCATGGCAAAGTAGGCCGACCAAGTCCGTCTACCTCGACCGC  
CGCAACAACGTCTAAAGCTCTCCAGGGCGAGTTCATGCCGTTGCGAAGCTCGAGGCC  
GTGTTCGGCACAGCCCGCTGGTCGACAGATCTTCATCTACGGCAACAGTGCCCGGGCC  
TACCCGCTGGCGGTGGTGTCCCGTCCGGGACGCCGCTTCTGCCATGGCATCGAGAAT  
CTCAAGCCCGTGTACCGGAGTCCCTGAGGAGGTAGCGAGGGCGCCGGCTGCAATCC  
TACGAGATTCCACCGCAGTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCTA  
CTCACCGGCATCCGCAAGCTGGCACGCCAGTGTGAAGAAGTTCTATGGCAACGTCTC  
GAGCGGCTCTACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCCGAGCTGCCGAA  
AGCGGTCCCGATGCCGGGTGCTTCCGACGCCGCTGTGCCGCGCTGCCGTTGCTGGG  
TCTACCGCTGCCGGATGTGCCGGCGACGCCGACTTCGCCGACCTGGTGGTGAACGCTC  
TCGGCGCTGCGTGGCAACCTGCTGCACGAGATCTCGCGTGCACGTGCCGGTGGT  
GTCATTGTCAGGCCGGCAAGCGACTGCCGACCATCGAAGCAGCGC  
ACCGCGTCAGGCCGACCCAGCTGCCCTCGATACACGGTCGCTCCGCACGGAAGTGCAC  
GCCAGCGACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCGCCGAGCCCCGAAC

**FIG. 8HHH**

CTGCCGGCACCAGCGCCAAGTGCACCGTACTGCTGACCGGCCACCGGCTTTG  
 GGTCGCTACCTGGCGTGGATGGCTGACCGCATGGACCTGGTCAACGGCAAGCTGATC  
 TGCCTGGTCCGCCAGATCCGACGGAGGAAGCACAAGCCGGCTGGACGCGACGTTGAT  
 AGCGGCAGCCGTATTGGTGCAGCACTACCGCAATTGGCGCCGGCGCTCGAGGTG  
 CTCGCCGGGACAAGGGGAGGGCGACCTGGGCTGGACCGGGTACACCTGGCAGCGGCTA  
 GCCGACACGGTGGACCTGATCGTGACCCCGGGCCCTGGTCAACCACGTGCTGCCGTAT  
 AGCCAGCTGTCGGCCCAAACCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACC  
 GGCAAGCGCAAGCCATACTACACCTCGACGATGCCGTGGCGAGCAGATCCCACCG  
 GAGGCAGTTCACCGAGGACGCCGACATCCGGGCATCAGCCGACCCGAGGATCGACGAC  
 AGCTACGCCAACGGCTACCGCAACAGCAAGTGGCCGGGAGGTGCTGCTGCCGAAGCT  
 CACGAGCAGTGGGCTGGCGTGAAGGTCTCCGCTGCGACATGATCCTGGCGACACC  
 AGCTATACCGGTCAACCTGGGACATGTTACCCGGCTGATGCTGAGCCTGGCC  
 GCTACCGGCATCGCACCCGGTTCTATGAGCTGGATGCGCACCGCAATGGCAACGC  
 GCCCACTATGACGGCTTGCCGGTCAATTGTCAGAAAGCATTGACCCCTGGGACA  
 CATAGCCGGACCGTTTCACCTACCGTGAACCCCTACGACGACGGCATCGGG  
 CTGGACGAGTCGTCGACTGGCTCAACTCCCCAAGTAGCGGGTCCGGTTGCACGATCCAG  
 CGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTGAGACTTCGCTGCGTGCCTGCCG  
 GATCGCCAGGCCACACCTCGCTGCTGCCCTGCTGCACAACCTACCGAGAGCCTGCAAAG  
 CCGATATGCCGGTCAATCGGCCAACCGACCAGTCCCGCTGCCGCAAGAAGCGAAA  
 ATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCAAGTACATCAGC  
 AACCTGCGACTGCTGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO.:86)**

>uniprot|Q7TY99|Q7TY99\_MYCBO PROBABLE FATTY-ACID-CoA LIGASE FADD9  
 (FATTY-ACID-COA SYNTHETASE) (FATTY-ACID-COA SYNTHASE)

MSINDQLTRRVEDLYASDAQFAAASPNEAITQAIQDQPGVALPQLIRMVM  
 EGYADRPALGQRALRFVTDPSGRTMVELLPRFETITYRELWARAGTLAT  
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL  
 RPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVFDYHGKVDTHREAV  
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG  
 APKGAMYRESQVMSFWRKSSGWFEPSGYPSTLNFMPPMSHVGGRQVLYGT  
 LSNGGTAYYVAKSDLSTLFEDLALVRPTELCFVPRIMVFAEFHSEVDR  
 RLVDGADRAALEAQVKELRENVILGGRFVMALTGSAPISEMATAWVESLL  
 ADVHLVEGYGSTSTEAGMVLDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR  
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMAKVGPQDFVYLD  
 RNNVLKLSQGEFIAVSKLEAVFGDPLVRQIIFIYGN SARAYPLAVVPSG  
 DALSRHGIENLKPVISESLOEVARAAGLQS YEIPRDFIIEETTPFTLENL  
 LTGIRKLARPQLKKFYGERLERLYTELADSQSNELRELRQSGPDAPVLPT  
 LCRAAAALLGSTAADVRPDAHFADLGGDSLSALS LANLHEIFGVDPVG  
 VIVSPASDLRALADHIEAARTGVRRPSFASI HGRSATEVHASDLTLDFKI  
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDRMDLVNGKLI  
 CLVRARSDEEAQARLDATFDSDGPYLVRHYREL GAGRLEVLAGDKGEADL  
 GLDRV TWQRLADTVDLIVDPAALVNHL PSQLFGPNAAGTAELLRLALT  
 GKRKP YIYTSTIAVGEQIPPEAFTEDADIRAI SPTRRIDDSYANGYANSK  
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA  
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAIC TLGTHSPDRFTYH  
 VMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALP  
 DRQRHTSLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPH  
 TAAIIAKYISNLRLGLL

**FIG. 8III****Q9CCT4****Nucleotide sequence (SEQ ID NO:87)**

>uniprot|Q9CCT4|Q9CCT4\_MYCLE Putative acyl-CoA synthetase

GTGTGGGCCAACAGTCTATTCTCATCGCAAGGAATCTGTTATGTCGACTATCACTAAG  
CAGGAAAAGCAGCTCGCACGCCCGTGTGACGACCTCACCGCCAACGACCCGCAGTCGCC  
GCCGCCAACCCGACCCGGCGTAGCCGCCCTTGCCCAGCCCCGGCTTCGACTGCC  
CAAATCATCCAGACCGCGCTGGACGGTTACCGGGAGCGGCCGACTGGGCCAGCGCTC  
GCCGAGTCACCAAAGACCTAACGACCGAGCGCACCTCGATGGAGCTGCTCCCCAGCTT  
GAGACCATCACCTACCGCCAGTTGGCGACCGTGTGGAGCGCTGGCGCCCTGGAGG  
CACGACCTACTGCACGCCGGTACCGGGTCTGCGTGTAGGTTCAACAGTGTGATTA  
GCCATCATCGACATGGCGCTGGCGTGTGGTGCTGTGGCGGGTCCACTGCAGACCA  
GCGGCGATCACCCAGCTGCAGTCGATCGACCGAGACCGAACCCAGTATGATCGCGAC  
AGCGTAAACCAGCTGCCGATACTGTCGAGCTGATCCTGTCTGGCCAGGCAGCGAAG  
CTCGTTGTGTTGACTACCACCCGAGGTCGACGAGCAGCATGACGAGTGGCAACCGCC  
CGGGCGCGGTTGGCGGACAGTAGCGTGGTGAGAGCCTGACCGAGGTTCTCGGTG  
GGCAAGACGCTGCCAGTACGCCATCCCCGTGGCGATGACTCTGCTGACCCGTTGGC  
TTGCTGATCTACACATCTGGCAGCACCGCGCACCAAGGGCGGATGATGTATCTGCAA  
AATGTCGCGAACAGATGTGGCGCCGGTCAGACGAAACTGGTTCGGGCAACCGCCG  
ATCACTCTAACCTCATGCCGATGAGCCACGTATGGGCCGCGGAATCCTCTACGG  
CTCGGTAACGGCGGACGGCTACTTCGCCGCCGAGCGACCTCTGACGCTGCTGGAG  
GATCTCAAGCTGGTGCACCGAGTTGACCGTCGGTTAGCCAACAGCGGCTCCGCC  
TACGATGAATCCAACCGCAGTTGACCGTCGGTTAGCCAACAGCGGCTCCGCC  
GCAGCCATCAAAGCGAACGTTATGGATGAAACAGCGGCAATCCCTGCTGGGAGG  
ACGGTACATCGCGGCTATGACGGGCTCGGCACACCTCCCCGGAGTTGAAACACGG  
CTACTCGAAATGCATCTGTTGAAAGGCTACGGCTCCACCGAACGGCGATGGTCT  
GACGGCGAAGTGCACGTCGCCGGTTATCGATTACAAGCTGGTCGACGTTCCGG  
GGCTACTTCAGCACCGACCGCTTATCCGAGAGGTGAAATTGCTGCTCAAGACCC  
GAAACATCGCGGCTACTACAAGCGTCCTGAGGTTACCGCCACCGTGTGACAGCG  
TACTACCAGACCGGAGACATTGTCGCCGAAGTCGGTCCGACCGGCTCGTACGT  
CGCCGCAACAACGTGCTGAAACTCGCGCAGGGCAGTCGTACCGTCGCGAAACTCG  
GCAGCGTCAGCAATAGCCCACGGTCCGGCAGATCTACATCTATGGCAACAGCG  
CCCTACCTGTTGGCTGTTGGCGACCGAGGATGCGTGGCTACCAATGACATTGAG  
GTGCTCAAACCGCTGATTATCGATTCTTACAGAAAGTAGCGAAAGAGCGACCTG  
TCCTACGAGGTGCCCGCGACTTAATCGTGAGACTACACCGTTCAGCCTGGAGA  
CTGCTACCGGTATTGCAAGCTGGCGTGGCGAACGCTCAAGCAGCACTACGGCG  
CTCGAACAGCTACGCCGATCTGGTTGAAAGGTGAGGCAAATGCACTGACGTG  
CAAAGCGTGGCGAACGCTCCGGTACTGCAAGACGGTGAGCCGAGCGTGGC  
GGAGTGGCGACCGATTGCCGTCGAATGCGCACCTCACCGACTTAGGAGGAGACT  
TTGTCGCGCTGACATTGGTAGCCTGCTACCGAACACTCTCGACATCGATGT  
GGCGTCATTGTCAGCCCTGTCAACAAACTGGTGGCGATGCCGACTACATCG  
CGGCAGGGCACGAAGCGGCCACTTCATTGCCATACACGGTGTGACGCTGG  
CATGCCAGTGACCTCACTCTAGACAAATTGATCGATGTACCGCTGACTGCC  
GTATTGGCGAACCCGGCACCGAGGTGCGCACCGTCTGGTACCGGGCGTACCG  
CTGGGGCGCTACTTGGCCCTGAAATGGCTCGAACGGATGGACCTGGTCAAG  
ATCGCTCGGTAAGAGCCAAGTCCAACGAGGACGCTGGGCCGCTCGACAAGAC  
GATAGCGAGACCCAAACTGCTGGCGCACTACCGAGGAACGGCAACCGACC  
GTCATCGCCGGCGACAAAGGCGAAGTAGATCTGAAATTGGACCGGAAAC  
CTGGCCGACACGGTCGATCTGATCGTGACCCGCCCTGGTCAACCACGT  
TACAGCGAGCTATTGCCCAATACGTTAGGACCGCCGAGCTGATTG  
ACCAAGTAAGCAAAGCCGTACATCTATGTGTCGACAATCGCGTCGG  
TAATCAGATTGAG

**FIG. 8JJJ**

CCAGCAAAATTACCGAAGACTCCGACATCCGAGTCATTAGCCCGACGCCAACATCAAC  
 ACAACTATGCCAACGGCTACGGCAACAGCAAGTGGCCGGCGAAGTGTGCTGCGCGAA  
 GCTCACGACCTATGCCGTCTGCCGTACGGCTTCCGCTGCACATGATCTGGCCGAC  
 ACCAGCTATGCCGTACGCTAACGTCCCCACATGTTACTCGAATGATGCTGAGCTA  
 GCCGCCACCGGCATCGCACCCGGCTCGTACAGCTAGACGCCGAGAGCAATCGGCAA  
 CGGCCACTACGACGGCTGCCGTGAGTTACGCCGAAGCGATCTCACCGAC  
 GACAAAGCCTGCACGATCGAGACGGGTCACGACCTATCATGTAATGAACCCGAC  
 GACGGCATCGGTATGGACGAGTTGTGGACTGGTTAATTGATGCCGCTGCCCTATA  
 CGCATCAACGACTACGACGAATGGCTGCGACGGTTGAGATTGCTGCGGCCCTGCC  
 GAAAGGCAGCGTCACAGCTACTGTTGCCGTGTTGACAACATACAGAACGGGAGAAG  
 CCATTGCACGGGTGCTGGCACCAATCCGGTCCGTACGGCGTTCAAAACGCGAAC  
 ATTGGTCAGGACAAAGATATTCCGCATATCTGCCGCAATCATGCCAAATATGTCAGC  
 GATCTGCAGCTGCCGCTGGTTGA

**Amino acid sequence (SEQ ID NO:88)**

>uniprot|Q9CCT4|Q9CCT4\_MYCLE Putative acyl-CoA synthetase  
 VWRQQSISHRKESVMSTITKQEQLARRVDDL TANDPQFAAKPDPAVAA  
 ALAQPGRLRPQIQTALDGYAERP ALQQRVAEFTKDPKTGR TSMELLPSF  
 ETITYRQLGDRVGALARAWRHDLLHAGYRVCVLGFNSVDYAIIDMALGVI  
 GAVAVPLQTSAAITQQLQSIVTETEPMSIATSVNQLPDTVELILSGQAPAK  
 LVVFDYHPEVDEQHDAVATARARLADSSVVVESLTVLGRGKTLPATPI P  
 VADDSDAPLALLIYTSGSTGAPKGAMYLQSNVGKMWRRSDGNWFGPTAAS  
 ITLNFMPPSHVMGRGILYGTLGNGGTAYFAARSDLSTLLEDLKLV RPTEL  
 NFVPRIWETLYDESKRAVDRRLANSGSADRAAIKAEVMD EQRQSLGGRY  
 IAAMTGSAPTSPELKHGVESLLEMHLLEG YGSTEAGMVLFDGEVQRPPVI  
 DYKLVDVPDLGYFSTDQPYPRGELL LKTQNMFPGYYKRPEVTATVFDSDG  
 YYQTGDIVAEVGPDRLVYD RNNVLKLAQGQFVTVA KLEAFNSPLVR  
 QIYIYGNSAH PYLLAVVVPTEDALATNDIEVLKPLI IDSLQKVAK EADLQ  
 SYEVPRDLIVETTPFSLENGLTGIRKLAWPKLQHYGARLEQLYADLVE  
 GQANALHVLKQSVANAPVLQTV SRAVGTILGVATTDLPSNAHFTDLGGDS  
 LSALTFGSLLRELFDIDVPVGIVSPVNVL VAIADYIERERQGTKRPTFI  
 AIHGRDAGKVHASDLTLDKFIDVSTLTAAPVLAQPGTEVRTVLLTGATGF  
 LGRYLALKWLERMDLIVEKGKIALVRAKS NEDARARL DKTFDSGDPKLLAH  
 YQELATDHLEVIAGDKGEVDLELDRQ TWRLADTV DLI VDPAALVNHVLP  
 YSELFGPNTLGTAE LIRIALTSKQKP YIYVSTIGVGNQIEPAKFTEDSDI  
 RVISPTRNINNNYANGYGN SKWAGEVLLREAH DLCGLPVTVFRCDMILAD  
 TSYAGQLNVPDMFTRMMLS LAATGIA PGSFYELDAESNRQRAHYDGLPVE  
 FIAEAISTLGDQSLH DRDGFTTYHVMNP H DGDGIGMDEFV DWLIDAGCPIQ  
 RINDYDEWLRRFEISL RALPERQRHSSLLPLLHNYQKPEKPLHGSLAPTI  
 RFRTAVQNANIGQDKDIP HISPAII AKYVSDLQLLGLV

**Q54JK0****Nucleotide sequence (SEQ ID NO:89)**

>uniprot|Q54JK0|Q54JK0\_DICDI Putative uncharacterized protein  
 ATGTTAAAACATATTAAAAATT TTTAAC TAGAAAAGAAGAAAGAAAAGAAAAGAAGTA  
 GAGAAATTAAAAGATGGAGTATCAATAACTGAGGTAAACAATCGAACCTAGTAGTTAT  
 TCATGCAATGGTTGGATCAGAGATATGCCACCAAAACAAGAGAGATATGCATGTAAT  
 GAATGCTAAATT CGATTATGAGTAGTGAGTGTATAGAAAAGAAATGATATTAATAAAT

**FIG. 8KKK**

GGTACACAAGAAGAGAAAGATAAATTAGTTAGTAGGGGAGAGTAATAATGGAATAAAGTAC  
GAGCCAGTGAGACATTATGATCCATCACCATACCTCATCAATTAAACATTAGAGAATGAG  
ACTCAATTCAATTAGTTAGTTACGTGGAATTCAACATTGAAACAATGGAGAAA  
TCATTTAAATACTTTAAAAACAGACCATGTCTGGTATTAGAGAGAGATTAGGAGAGGAT  
AATGTTTATCAGAAAGATATAATGGTAACATATGGTGAAGTGTATGAGAAAATCTTA  
ACCTGGCAAAGGCATTAACTAATTTCATCGAAAGAAGAGATTTCATTCAATCTATATG  
GATAATTGCTTAGAATGGTATTTCACAGATTTGCATCATTATGGGCCGATTAAATAGTG  
GTACCATTACATCATGCTCAAATAGTTAATCTTAGAGATTCTTGGAAATTCCGAA  
TCAAAATGTATAGTTGTTCTGGTGAATCATTAAAAATTAAATAGAAACTTTATGATCAA  
TTGACTGAGCAAGATAAATTAGAGAACCAATAGTGTGAAATGATAGTCATAAGGAG  
GATCTATCGATCAGTCATTAGTCGATAGATTCCAAGTGGCGTAGAATTAAAACCTTC  
AATGAGATGATTAAAATTGGGAATCATTAAGTCAGGCTAAATATGAATTGTCCCAGTT  
GGTCAAATGATCTTCCTCGGTGACTTACAAAGTGGTAGTACTGGTGTACCAAAGGT  
GTAATGAAGTTAGATTCAATTTCATTACTAATTGTCATTCTATGTCATTCCCA  
AATGCAGTTATAGTTATAATACCCATCACATAGTCACGTTAAGTGATTGGAGATAT  
ATTATATGGGTGGTAGAGTAGCTATCTATTAGGTGATATGAATCTATTATGAAAGAT  
TTAGCCTGGTTAGACCTCATTCAATTGGGCTGTACCAAGATTGGAATTATTATTC  
ACCCAGTCAAGAGTGATCTAAAGCAATACATGTTGAAAATCCACAATTGGATGAAAGA  
ACTGCCACACTCTATTGCTATAAAGGTATTAGAAAGTTATTAGGTGATAGAAATTAAAT  
CTAGTACTGGTGGTCCGACTGCAAATGAAGTACTCAAATTATGAGTGATTGTGG  
AAAGATATAAACATTCAAAATTCTATGGTTAACTGAAGTATCAGGTGTTGATAGAT  
GGTTATATCTGACGAAGTAGAATTCAAATTGAAACCGTACCTTCTTGGATATTAC  
CCAAC TGATTACACATCCTCGGTGAATTGGTTGTAATTACATCAACAATGTCGCA  
GGTTATTATAAAACTCAATTAACTCAGAATCATTGAAGGTGGTTAAAAC  
GGTGATGTCGTAGAATTAAATTGGAGTTAGAAAAGTTAAACATTGATAGAAATTAAACAT  
GCCTTCAAATTGGCAAATGGAGAATTGTTACACCAGAACCATGGAAAATAATTGTT  
TCACTTGTATTAACTCAAATTTTATTGTAATTCACTTAAACATTTTAGTTGCA  
ATTGTTAAACCACATCACAAGATTGTTAAAACAATTAGGACTTCAGATATACCAATCGAT  
CAATTAAATTGAAAATCCAACCTTAAACATTCAAACACTTTATCAGAGATTAATAAATTCA  
AAAGAAAAAAACTAGCAAATTGAAATTCCAACAAATTACAAATAGATTCACTGAA  
TGGACAATTGATAATAATTAACTGGTTCTGGTAAATTAAATAGAGGTGAATTATAT  
AAATTATATAAAATTAAATAATATGTTGATATAATTGATAAAATTCAACAAGGT  
TTAAGAAATAATAATAATAATAATTGATAATTAAATAATAATTGATAATAATAAT  
AATAATGAATCAAATAAGATAATTGAAATTATATAAAATCAATTAAATTAGAC  
GGTAGAATTGAAGATAATAATTGAAATTGAGAATTATCATTATTCAATTGGTGGT  
GATTGTTAGGTGCTGTTAAATTATCATCACTTTAAAAGAAAAAGAAAATTGATATT  
TCACCTTCAACAATTAAATCAAATTAAATTATCAGAGATTTATCTTCAATTGAAATTAAAT  
GAAAAAGAATCAAATCAATTGTTGAAGATTAAAGAAAATTAAATCAATTG  
AATGAAGAGATGATTAGATGAAGATATTAAAAATCAATTGACCAAATTAAAATGCA  
CAACCATCATCAACTCCCTCTTCATCAAACACCATCACATCATCATCACCA  
CCACCATCATCAAATTCAAATAATTGGTCAAATTGCTTTCATATGAAATCAATT  
ATTACAGGTGTTACAGGTATTAGGTACATTAAATTAAATTAGAGGATAAA  
TCAATTGGTATTGAGAGAATTATTGTTAGTTAGAAATGTAAGAATGAAGAAGAGT  
TTAAATTAAATTGAAAGAATTGAAAATCTTGTATCAATTGATGAATGAAAAGATT  
AGAGAAAAGGTAACTTCAGTTGTGGTGAATTCAAACACATTTCGGTGGTTCTACT  
GAAACCTCAAATGTTATCTTAGCGGTGATATGTAATTCAACATGGGCCATTGTT  
AATATGGCCTATCCATATGCGAATATGAAATCAACAAATGTTACATCAACTCGTGATATC  
CTAAGATTATGCACTACCGGAAGAGGCCTCTTAAAGTTGGTCTACGTTCAACAGTT  
GGTGTATTCTTGGAAATGGTGTGAAAGATAGATGAATCAACAGCACCACAACTTC  
TTTTAGATCATGGTAAATTGTTAAACCAACAAACTAATATCAGATATACTGTTAGA  
GAAGCGGCTTCATATGGTTACCAACAATTGTTAGTTCAACATGGTAAATTAAAGGTATCCTA  
ACCAACATCTGGTTCAACAATCAAATGATTCAATTGTTAAATAATTAAAGGTATCCTA  
GGTCGAGTTCTATCCAACAAAAAGATTACTCTAGTGGCATTAAATCTTCACCA  
GTTGATTGGGTATCATCTTCATGGTTCTTAATTAAAGCATCTTCATTGGTGTAAAT

**FIG. 8LLL**

AATAACAAAATTATCATATGGTAAATGATAATCGTTATCTTAGATTGCTATGTCAA  
 TATATAAATAAGAAAAACAATTAGAAGAAATTAAATTATTCGATTGGATCGATGCTCAA  
 CTTAATTCTTCAAATAATCCATTGTATTCTATTAAACATTATTAAAAAGAATGATCGT  
 TTCCAATTGGTTCTCAGTCAATTAAAATCCAAAACATTAAAGATTAGAATCAATT  
 GGTGAACCTCAATGTCAACCAATATCTGATTCCACAGTAATCAATTATGTAAAATATTAA  
 ATCTCAAATAATTAAATTCAAACAATTAAATAAAATAA

**Amino acid sequence (SEQ ID NO:90)**

>uniprot|Q54JK0|Q54JK0\_DICDI Putative uncharacterized protein

MLKHKNFLTRKEERKEVEKLKDGVSITEVKQSNLVVYSCNGCGSEIWPPKQERYACNECSN  
 FDLCSECYRKEMILINGTQEEDKLVSGESNNNGIKYEPVRYDPSPLPHQLTLENETQFQLVYSL  
 RGNSTFETMEKSFKYFKNRPLGIRERLGEDNVLSERYKWLTYGEVYEKSLLAKALTNFIERRD  
 FISIYMDNCLEWYFTDFASLWAGLIVPLHHASNFSNLLEILWNSESKIVCSGESFKNLIELYDQL  
 TEQDKLEKPIVLKLIVHKEDLFQSLVDRLPLSGVEFKTFNEMIKIGESLSQAKYEFVPVGPNDLSS  
 VTYTSGSTGVPKGMKLDISFNLLIVNSYVQFPNAVSYNTLSHSQRSLSDWRYIYMGGRAVAYSG  
 DMNLLFEDLALVRPHSFWAVPRFWNLLFTQFKSDLKQYMFENPQLDERATLYCYKGIRKLLGD  
 RINNLVTGGAPTANEVLKFMSDCWKDINISNSYGLTEVSGVCIDGYISDEVEFKIEPVPSFGYYPT  
 DLPHPRGELVVKSSTSAGYYKNTQLTSESFEFGWFKTDGVVELIGVRKVLIIDRIKHAFKLANG  
 EFVTPPEPLENNFVSLCINQIFIYGNSLKTFLVAIVKPSQDCLKQLGLQDIPIDQLENPTLKSLLSEI  
 NKISKEKKLANYEIPKIITIDFTEWTIDNKLTGSGKFNRGELYKFYKIKINNMFDIIDKIQQGLRNNNN  
 NNNNDNINNNDNNNNNESNKNDNFENYIKSILNLDGRIEDNNFNLENLSFIQIGGDSLGAVKLSSLL  
 KEKENIDISPSTILNQNFnLSSLKLINEKESENQSIVEDFKENFKINWNEEMILDEDIKKSIDQIKNAQ  
 PSSTPSSSKSTPSQSSSPPSLNSNNIGQNAFHKMSIFITGVTGYLGTFLLFNLEDKSIGIERIY  
 CLVRNVKNEEEGFKLERIFEKSCINGMNEKIREKVI PVCGLSKPFFGVSTETFKMLSLAVDMVIH  
 NGAIVNMAVPYANMKSTNVTSTRDILRLCTTGRASFKKLVYVSTGVFGNGDEKIDESTAPSTF  
 FLDHGNGYNQTKLISDILVREAASYGLPTMIFRPGTIFSHTQSGFNNQNDSIGLIKGILGSSSYPTK  
 KDYSSGDLNLSPLDVWSSSMVSLIKHLPFWCNNTKIHMVNDNRLSLDLLCQYINKEKQLEEINY  
 FDWIDAQLNNSNNPLYSIKHLFKKNDRFPIGSQSIIKNPKTIKDLESIGELQCQPISDSTVINYVKLI  
 SNNLIQTINK

**Q2MFQ3****Nucleotide sequence (SEQ ID NO:91)**

>uniprot|Q2MFQ3|Q2MFQ3\_STRRY Putative non-ribosomal peptide synthetase

ATGACCGACACGTACGTTCTCACGCCGTTATCCAAGCGGCCCCAGGTCCCGGGTGCC  
 CGCACGCCGGCGCCCGCTACCCACGGGACAGCCGCATCCCGGAGCTGTTGAGGCACAG  
 GCCGCGCGCTGCCGCAGGCCCCGGGGCCGGCACGGGACCCGACCTGACCTACGGC  
 CAACTCGACGCCACGCCGACCGTGGCGGACCCGCTGGCTGCCGGGGGTCCGGCG  
 GGTGACCTGATCGGCGTGTGGCGCAGCCGTTCCCTGGAGGGCGCTGGTGGCGCTGCTGGGC  
 ATCCTCAAGGCCGGCTGCCGTACGTACCGCTCGACGAGGAACGTGCCGGGGCTG  
 CGGGCCATGGCCGAGGACGCCGCATAGCGCCGGTCACCTGCCGGCAGCACCGC  
 CGGGTGCAGGGCTACCGCGTCAGTCAGGTCGGCTCCCTGGCCGGGGCGAG  
 CGCGCGAGCGCCCCGCCGACCGGCCACCGGCTCCGCCGACTGCGCTACGTC  
 GCCTTCACTTCCGGCACGCCGGCCAAGCCGTTAGCGCTGTCCCACCGGGCGT  
 GTCCGCCTCGTGTCCGACCCGGCTCACGCCACCGGGACCGGGGAGGTGCTG  
 CACGCCCTACAGCCTGCTCCGACGCCCTCGACCATCGAGATCTGGGGCGCCTGCTGACC  
 GGCCTGCTGGCTGGCGCAGCGCAGGAACGTGCTCGCCCACCGCCCTGGAACGG  
 CTGCTCCGCGCGGGCGCTACCGTGGCGTACCTGACGACGAGCGTCTCCACCTCGTC  
 GCGCGGACCCGCCCGAGGCGCTGGCCGGCTGCGGTTCGTCTCCGCGGGGGAGGCG

**FIG. 8MMM**

ATGGACCCGCCCTGGCGAACGCCGCTCCTCGGCCCTGCCCCCGACCACGGTGGTCAAC  
 TTCTACGGCCCGACCGAGAACGCCGTTCTCCACCGCCCATGTGCTCACCCCCCTCCCC  
 GAGGACGCCGACACGTCCTGGGACGCCGCTGGGACGCCGCTTCGGCGCTTCCACCTGCCACGTCTG  
 CGGGCCGACGGCTCGCCCCGGCCGGGGAGGAAGGGGAGCTGTACGTGGCGGGGAC  
 GGGCTGGCGCTGGGCTACCTCGGCGACGCCGCTGGGACGCCGAGCGGGTCTGTGACGCTG  
 CCCCGGGTCGAGCCGGACGGACCGCTGTACCGGACCCGGGACCGGGGCTACGGCACGCC  
 GACGGGCTGCTGGAGTACCGCGGACGGCTCGACCGCCAGGTCAAGCTGCGCGGCCGCC  
 ATCGAGCTGGACGAGGTGAGAACCGCTAACGGCCCACCCGAGGTGGCGAAGCGGCC  
 GTCGAGGTGACGGGACTCCCTGACCGCTAACGTACGGCACCGTCCCCGGCGGCC  
 CTGCGCTGGCCGACCTGCGCGCTACTGCGCCAAGTGGCTGCCCCCGCAGGGCGTCCC  
 GCCCTGATAACCCCTGGACCGCTCCCGGTCACCAAGCGGGCAAGATGACCGCAGCC  
 CTGAGGCCGACCGCGGCCACCGCCGGGGGGGAAGACACCGCGGGAGGGCCGCGGCC  
 GACGAGCCGGAGGCCACCGACGGCTGTCCGGTCTCTCGGAAGTGTGGCACCGAGTG  
 CTGCGTGTCCGGCCCACGCCCGGGACGACTTCTTCTCTCGCGGGGACTCCCTGCTC  
 GCCTCGGAGACCGTCACCCGCACCCCTGCCGTACTGGCCTCGACGCCCTGGCTCC  
 ACCCTCATCAGGGCGCTGCTGGCGGCCACCCCGAAAGACACCGCGGGAGGGCCGCGGCC  
 GGAGTCCGGCGGCCACCGCGGCCAGGGCCGGCCAGGAACCGCCGTCGACTTCGCC  
 GCGGAGACCGGACTCGGCTTCGCCCTCCCGCCCGAAGGCCCGGCCGAACCCGAC  
 GACCCCGAGGACGTCCTGCTCACGGCGCTCCGGCTCGTCGGCGATTCTGTCAC  
 CGTCTGCTGCACCGACGCCGCCCGTACGCCGACTGCCGCTACGGCGACGAGCCCGC  
 CACGCCCGCAGCGGGTCCGACCGCCCTCACCGCTACGGCTGACCTCGACGAGGCC  
 GACTGGCAGCGCGTGGAGTGCTTCCCGGGGACTGACCCAGCCGCGCTGGGCTCGAC  
 CACCGAGCGCCGACGCACTGGCCAGCGCCTGACCTGATCGTCACAACGGTGCCCG  
 GTCAACTTCCCTAACCCCTACCAGCAGTTGCGCCCGCGAACGTCGACCGAACCCCGAG  
 GTCGTCCGGATCGCCGCGCCGGTGCCTGCGACTTCGTGTCACCCTCGCAGTC  
 GTCGCGGGCTTCGGCACCGCCGGGGTGCAGGGTGGACGAAAGATCTGCCGCC  
 GCGACGGGCTGACCATGGGTACCGGAGAGCAAGTGGTCGCCGAAGGGGTGCTCGG  
 CAGCGGCCGCGCAGGGCTGCCGGTGGCGTGTACCGGCCGTACGAGGTACGGCGAC  
 CGGACGCACGGCGCGTCAACACCGAGACGCCATCTGCTCGCTGTTCAAGATGATCGCC  
 GACACGGGAGTGGCCCCGACATCAAGCTGCCGATGGACTTCGTACCCGTCACCACCTC  
 GCGAGTCCCTGGTGCACATGCCACCGCCGCCGACGCCGGGTCTACCACCTG  
 ACCAACCCGCCGCCGGCGATGCTGTCGGACGTCCTCGACCGGATGCGCGCCGGCTTC  
 ACCCTGCGCACCTGCCGTACGACGCCGTGGTCCGCGAGCTCGTCCGGCACGTCGCGAG  
 AACCGACGAGGCCACGCCGCTCGTGTCCCTGTGCGTGGACCGCAGCCGACCGCC  
 GACATGTCGTCAGAGATGTACCTCAAGGGCACCTCCGGTCTGGGCGGCCAAC  
 GCGAGGAGGGCTGGCCGAGCGTCACTGCCGCGGCTGACTGCCGCCGTGACTCCGCTGCTG  
 GACCGCTACCTGGAGTACTCTCACCTCCGGTACCTCACCGCCGCCGGCGCC  
 GGGCCGAATCCGAGGCCAGCGAATACCGAGGACGAGCCGGTGTCCGGGACCGAACCG  
 ATATCCGGGACCGAACCAGATTCTGCCGCGAGCCGATATCCGGGACCGAACCGATTCT  
 GCGCCGGGCCGATATCCGGGACCGAGCCGATACCGCCGCCGAGCCGATATCCGGGACC  
 GCAGCCGCAGCCCGCACGGAGCGCAGCCGATGA

**Amino acid sequence (SEQ ID NO:92)**

>uniprot|Q2MFQ3|Q2MFQ3\_STRRY Putative non-ribosomal peptide synthetase

MTDTYSSRPLSKRPQVPGARTPAPGYPRDSRIPELFEAQAAALPQAPAAHRGDRTLTYGQLDAHADALAD  
 RLAAGGVPGDLIGVCGSRSLAEVALLGILKAGCAYVPLDEELPPARLRAMAEDAGISAAVTLPGSTRRV  
 RGLRVSVEVGSLGRPAPERASGPAPDRATGSAADCAYVAFTSGTGRPKPVALSHRGVVRVLVSDPGLTPP  
 GPGDGVHLAYSLSSEDASTIEIWGALLTGACLVVADREELSPTALERLLRAGGVTVAYLTTSVFHLVARTR  
 PEALAGLRFVSAGGEAMDPRLANAVLAACPRTTVNFYGPTENAVSTAHVLTPLPEDAAHVPLGRPFgas  
 TCHVLRADGSPARPGEEGELEYVGGDGLALGYLGDPQLTAERFVTLPAVEPDGPLYRTGDRAVRHADGLLEY  
 RGRLDRQVKLRGARIELDEVETRLRAHPEVGEAAEVVDGHSLTAYTATVPGRPLPLADLRAYCAKWLPPQ  
 AVPALIPLDRFPVTSGGKIDRSRLKPTAPPPGPEDTAEAARRPDEPEATDGLSGLLSEVWHQVLRVRPTPR

**FIG. 8NNN**

DDFLLGGDSLLASETVRTLAVLGLDAALGSTIRALLAAPTLESFTAARVGVRGGTGGPAGGQEPAVDF  
AAETGLGFALPPAEGPAPNPHDPEVDLTLTASGFVGGFLHLRLLHATAARVHCPVRATSPAHARQRVRTAL  
TRYGLHLDEADWQRVECPGDLTQPRLLDHERADALAQRLLDILVHNGARVNFLYPYQQLRPANVDGTREV  
VRIAARRRVPHVVFVSTVAVVAGFGTAGVREVDEDLPPAHADGLTMGYAESKWVAEGVLRQAAAQGLPVAVY  
RPFYEVTDRTHGACNTETAICSLFKMIAADTGVAPDIDKLPMDFVPDFLAESLVHIATHRPADGRVYHTNP  
RPAMLSDVLDRMRAAGFTLRLPYDAWVGEVLRVAENPTSATAPFVSICVDRSRADMSVKEMYLKGTFP  
VLGRRNAEALAGSGLHCPPVDSALLDRYLEYFFTSGYLTRPAAGPGPESEAERIPEDEPVSGTEPIGS  
PISAAEPISGTEPISAAGPISGTEPIPAAEPISGTAARTERS

**ZP 03429464**

**Nucleotide sequence (SEQ ID NO:263)**

>gi|189214978:54695-58201 Mycobacterium tuberculosis EAS054  
NZ\_ABOV01000087, whole genome shotgun sequence

ATGTCGATCAACGATCAGCGACTGACACGCCGCTCGAGGACCTATACGCCAGCGACGCCAGTTGCCG  
CCGCCAGTCCCACAGAGGCATACCCAGGGCATCGACCAGCCCCGGGTCGCGCTTCCACAGCTCATCCG  
TATGGTCATGGAGGGCTACGCCGATCGGCCGCACTCGGCCAGCGTGCCTCCGCTCGTACCGACCC  
GACAGCGGCCGACCATGGTCGAGCTACTGCCGCGTTGAGACCATCACCTACCGCAACTGTGGGCC  
GCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCCGCATCGGCCGGGCGACCGGGTTGCGTGC  
GGGCTCAACAGCGTCGACTACACAACCATCGACATCGCGCTGATCCGGTTGGGCCCGTGTGGTTCCA  
CTGCAGACCAGTGCGCCGGTCAACGGGTTGCGCCGATCGTCACCGAGACCGAGGCCACGATGATGCCA  
CCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGCCGGTCACGCCCCGGCCGGCTGGTGTATT  
CGATTACACCGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCGGCCGAGCTCGGTTGGCCGGCTCG  
GTGACCATCGACACACTGCCGAACTGATCGAACCGGGCAGGGCGCTGCCGCCACACCCATTGCCGACA  
GCGCCGACGACGCCGCTGGCGCTGCTGACTTACACCTCGGTAGTACCGGCCACCCAAAGGCCATGTA  
TCGGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGCTGGTTGAGCCGAGCGGTTACCCCTCG  
ATCACGCTGAACCTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCACGGGACGCTTCCAACG  
GCGGTACCGCTACTCGTCGCCAAGAGCGACCTGTCGACGCTGTTGAGGACCTGCCCTGGTGCAGGC  
CACAGAATTGTCGTTCTGCCGCGCATCTGGGACATGGTGTGCAAGAGTCCACAGCGAGGTGACCGC  
CGCTGGTGGACGGCGCCGATCGAGCGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCCGAGAACGTGC  
TCGGCGGACGGTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAGATGACGGCGTGGTCGA  
GTCCTGCTGGCGACGTGCATTGGGAGGGTTACGGCTTACCGAGGCCGAGCTGGTCTGAACGAC  
GGCATGGTGCAGCGCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCGAGCTGGCTACTTCGGCA  
CCGATCAGCCCTACCCCCGGCGAGCTGCTGGTCAAGACGAAACCATGTTCCCGCTACTACCGCG  
CCCCGATGTCACCGCCGAGGTGTTGACCCGACGGCTTACCGGACCGGGACATCATGCCAAAGTA  
GGCCCCGACCAAGTCGCTACCTCGACCGCCGCAACAACGTCTAAAGCTCTCCAGGGCAGTTACCG  
CCGTGTCGAAGCTCGAGGCCGTTCCGGCAGAGCCGCTGGTCCGACAGATCTCATCTACGGCAACAG  
TGGCCGGGCTACCGCTGGCGTGGTTGTCCTCCGGGACCGCGCTTCTCGCATGGCATCGAGAAT  
CTCAAGCCCGTATCAGCGAGTCCCTGCAAGGGTAGCGAGGGCGCCGGCTGCAATCCTACGAGAATT  
CACCGCAGTTCATCATCGAAACACGCCGTTACCCCTGGAGAACGGCCTGCTCACCGGATCCGCAAGCT  
GGCACGCCCGCAGTTGAAGAAGTTCTATGGCAACGTCTCGAGCGCTCTACCGAGCTGGCGATAGC  
CAATCCAACGAGCTGCGCGAGCTGCGCAAAGCGGTCCCAGTCGCGCCGGTCTCCGACGCTGTGCCGTG  
CCCCGGCTGCCCTGCTGGCGTACCGCTGCCGATGTCGGGGGACCGGGACTTCCCGACCTGGCTGG  
TGACTCGCTCTCGCGCTGTCGTTGGCAACCTGCTGACGAGATCTTCGGCGTCGACGTGCCGGTGG  
GTCATTGTCAGCCCGCAAGCGACCTGCGGGGCTGGCGACCATCGAAGCGCGCACCGCGTCA  
GGCGACCCAGCTTCGCTCGATACACGGTCGCTCCCGACCGAAGTGCACGCCAGCGACCTCACGCTGGA  
CAAGTTCATCGACGCTGCCACCGTGGCGACGCCGAACCTGCCGGCACCGAGGCCAAGTGGCACC  
GTACTGCTGACCGCGCCACCGCTTTGGGTCGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACC  
TGGTCAACGGCAAGCTGATCTGCCTGGTCCGCGCAGATCCGACGAGGAAGCACAAGCCGGCTGGACGC  
GACGTTCGATAGCGCGACCGTATTGGTGCAGGCACTACCGCGAATTGGCGCCGCCGCTCGAGGTG

**FIG. 8000**

CTCGCCGGCGACAAGGGCAGGCCGACCTGGGCCTGGACC GGTCACCTGGCAGCGGCTAGCCGACACGG  
TGGACCTGATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTGGCCAAA  
CGCGCGGGCACCGCCGAGTTGCTTGGCTGGCGCTGACC GGCAAGCCATACATCTACACCTCG  
ACGATCGCCGTCGGCGAGCAGATCCCCCGGAGGGCGTTCACCGAGGACCCGACATCCGGGCCATCAGCC  
CGACCCGCAGGATCGACGACAGCTACGCCAACGGCTACCGA ACAGCAAGTGGGCCGGGAGGGTGCTGCT  
GCGCGAAGCTCACGAGCAGTGC GGCTGCCGTGACGGTCTCCGCTGCCGACATGATCCTGGCCGACACC  
AGCTATAACCGGTCACTGCCGGACATGTCACCCGGCTGATGCTGAGCCTGGCGCTACCGGCA  
TCGCACCGGTTGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACCGGCCACTATGACGGCTTGCC  
GGTGAATTGTCGCAGAACCCATTGCAACCTTGGGACACATAGCCGGACC GTTTGTCACCTACCAC  
GTGATGAACCCCCTACGACGACGGCATGGGCTGGACGAGTTGCTGACTGGCTCAACTCCCCAACTAGCG  
GGTCGGTTGCA CGATCCAGCGGATGCCGACTACGGCGAGTGGCTGCA CGGTTGAGACTTCGCTGCG  
TGCCTTGCCGGATGCCAGGCCACGCCCTGCTGCTGCCCTGCTGACA ACTACCGAGAGCCTGCAAAG  
CCGATATGCGGGTCAATCGGCCACCAGTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGG  
ACAAAGACATTCCGCACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGGCT  
GCTGTGA

**Amino acid sequence (SEQ ID NO:264)**

>gi|215431545|ref|ZP\_03429464.1| fatty-acid-CoA ligase fadD9  
[Mycobacterium tuberculosis EAS054]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAI DQPGVALPQLIRMVMEGYADRPALGQRALRFVTDP  
DSGRTMVELLPRFETITYRELWARAGTLATALSAPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVP  
LQTSAPVTGLRPIVTEPTMIATSIDNLGDAVEVLAGHAPARLVFDYHGKVDTHREAVEAARARLAGS  
VTIDT LAELIERGRALPATPIADSADDALALLTYTSGSTGAPKGAMYRESQVMSFRKSSGWFEPGYP  
ITLNFMPMSHVGGRQVLYGTLNGGTAYFVAKSDSLTLFEDLALVRPTELCFVPRPRIWDMVFAEFHSEVDR  
RLVDGADRAALEAQVKAELENVNLGVGRFVMA LTGSAPI SAEMTAWVESLLADVHLVEYGSTEAGMVLND  
GMVRRPAVIDYKLVDVPELGYFGTDQPYPRGE LLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDI  
GPKDFVYLDRRNNVLKLSQGEFIAVSKLEAVFGD SPLVRQIFIYGN SARAYPLAVVPSGDALSRHGIEN  
LKPVI SESLQEVARAAGI.QSYET.PRDPIIETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADS  
QSNELRELQSGPDAPVLP TL C RAA AL LGSTA ADV PDAH FADLGGDSLSAISI ANLIHEI FGVDVPVG  
VIVSPASDLR ALADHI EAARTGVRRPSFASIHGRSATEVHASDL TLDKFIDAA T LAAAPNLPAPSAQV  
VLLTGATGFLGRYLA LEWLDRMDLVNGKLICLVRARSDEEAQARLDATFD SGD PYLVRHYREL GAGRLEV  
LAGDKGEADLGLDRV TWQRLADTVLIVDPAALVN H VLPYSQ LFGPNAAGTAELLRLALTGKRKP YI YTS  
TIAVGEQIPPEAFTE DADIRAI SPT RRI DDSYANGYANSK WAGEVLLREAHEQCGLPVTVFRC  
DMILADT SYTGQLNLPDMFTRLMLSLAATGIA PGFSYELDAHGNRQRAHYDGLPVEFVAEAI CTLGTHSPDRFV  
YH VMNPYDDGIGLDEFV DWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALPDRQR HASLLPLLHN  
YREPAK PICGSIAPTDQFRAAVQEA KIGPDKDIPH LTAAIIAKYISNLRLLGILL

**ZP 03533123****Nucleotide sequence (SEQ ID NO:265)**

>gi|192384451:15291-18572 Mycobacterium tuberculosis GM 1503  
NZ\_ABQG01000169, whole genome shotgun sequence

ATGGTCGAGCTACTGCCGGGTTCGAGACCATCACCTACCGCGAAC TG GGGCCGGCCGGCACATTGG  
CCACCGCGTTGAGCGCTGAGCCCGCGATCCGGCGGCGACCGGGTTGCGTGCTGGCTTCAACAGCGT  
CGACTACACAACCATCGACATCGCGCTGATCCGGTTGGCGCCGTGCGTTCCACTGCAGACCA GTGCG  
CGGGTACCCGGGTTGCGCCCGATCGTACCGAGACCGAGCCGACGATGATCGCCACCAGCATCGACAATC

**FIG. 8PPP**

TTGGCGACGCCGTCGAAGTGTGGCCGGTACGCCCGGCCGGCTGGCGTATTGATTACACGGCAA  
 GGTGACACCCACCGCGAGGCCGTCGAAGCCGCCGAGCTGGCTGGCCGGCTGGTACCATCGACACA  
 CTTGCCGAACTGATCGAACCGCGCAGGGCGCTGCCGCCACACCCATTGCCGACAGCGCCGACGACGC  
 TGGCGCTGCTGATTACACCTCGGGTAGTACCGCGCACCCAAAGGCGCCATGTATCGCGAGAGGCCAGGT  
 GATGAGCTTCTGGCGCAAGTOGAGTGGCTGGCTGAGGCCGAGCGGGTACCCCTCGATCACGCTGAACCTC  
 ATGCCGATGAGCCACGTCGGGGCCGTCAGGTGCTCACGGGACGCTTCCAACGGCGGTACCGCCTACT  
 TCGTCGCAAGAGCGACCTGTCGACGCTGTCGAGGACCTCGCCCTGGTGCGGCCACAGAAATGTGCTT  
 CGTGCACGCGCATCTGGACATGGTGTTCGAGAGTTCCACAGCGAGGTGACCGCCGCTGGTGGACGG  
 GCCGATCGAGCGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCCGAGACGTGCTGCCGGACGGTTG  
 TCATGGCGCTGACCGGTTCCGCGCCATCTCCGCTGAGATGACGGCGTGGTCGAGTCCCTGCTGCCGA  
 CGTGCATTGGTGGAGGGTACGGCTCCACCGAGGCCGGATGGTCTGAACGACGGCATGGTGCACGCC  
 CCCCGGGTATCGACTACAAGCTGGTCGACGTGCCGAGCTGGCTACTTCGGCACCGATCAGCCCTACC  
 CCCGGGGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGC  
 CGAGGTGTTGACCCCGACGGCTTCTACCGGACGGGGACATCATGGCAAAGTAGGCCCCGACCAGTTC  
 GTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCATGCCGTGTCGAAGCTCG  
 AGCGGTGTTGCCGACAGCCCCTGGACAGATCTTACGCAACAGTGCCGGGCTACCC  
 GCTGGCGGTGGTGTCCCGTCCGGGACCGCGTTCTCGCCATGGCATCGAGAATCTCAAGCCGTGATC  
 AGCGAGTCCCTGCAGGAGTAGCGAGGGCGCCGGCTGCAATCCTACGAGATTCCACGCGACTTCATCA  
 TCGAAACCACGCCGTTACCCCTGGAGAACGGCTGCTACCCGACATCGCAAGCTGGCACGCCGAGTT  
 GAAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTACCGAGCTGGCGATAGCCAATCCAACGAGCTG  
 CGCGAGCTGCCGAAAGCGGTCCCAGTCGCCGGTCTCCGACGCTGTGCCGTGCCGGCTGCCGTG  
 TGGCTCTACCGCTGCCGATGTGCCGGACGCCACTTCGCCGACCTGGGTGGTGACTCGCTCTGCC  
 GCTGCGTGGCAACCTGCTGCCGACAGAGATCTCGCGTGCACGAGATGGCTGCGACCTGGGTGGTGCTATTGTCAGGCC  
 GCAAGCGACCTGCCGGCCCTGGCGACACATCGAACGAGCGCGCACCGCGTCAGGCACCCAGCTCG  
 CCTCGATAACCGGTCGCTCCGCGACCGAACCTGCCGGACCGAGCGCCAAAGTGCACCGTACTGCTGCC  
 TGCCACCCCTGCCGCGACCCCCGAAACCTGCCGGACCGAGCGCCAAAGTGCACCGTACTGCTGCC  
 GCCACCGGCTTTGGGTCGTAACCTGGCGCTGGAATGGCTGACCGCATGGACCTGGTCAACGGCAAGC  
 TGATCTGCCCTGGTCCCGCCAGATCCGACGAGGAAGCACAAGCCGGCTGGACGCCACGTTGATAGCG  
 CGACCCGTATTGGTGCAGGCAACTACCGCAATTGGCGCCGGCCCTGAGGGTGTCTGCCGGGACAAG  
 GGGCAGGCGCACCTGGGCTGGACGGGTCACCTGGCAGCGCTAGCCGACACGGTGGACCTGATCGTGG  
 ACCCCGCGGCOCTGGTCAACCACGTGCTGCCGTAGCCAGCTGTTGCCCAAACGCCGGGACCGC  
 CGAGTTGCTCGGCTGGCGCTGACCGCAAGCGAACGCCATACATCTACACCTCGACGATGCCGTGGC  
 GAGCAGATCCCGCCGGAGGCCTCACCGAGGACGCCACATCCGGCCATGCCCGACCCGAGGATCG  
 ACGACAGCTACGCAAACGGCTACCGAACAGCAAGTGGGCCGGCAGGTGCTGCGCGAACGCTCACGA  
 GCAGTGCAGGCTGCCGGTACGGTCTCCGCTGCCGACATGATCTGCCGACACCAGCTACCGTCAG  
 CTCAACCTGCCGGACATGGTACCCGGCTGATGCTGAGCCTGCCGCTACCGCATCGCACCGGTTG  
 TCTATGAGCTGGATGCGCACGGCAATGGCAACCGGCCACTATGACGGCTTGGCGTCAATTGTC  
 AGAACGCCATTGCAACCTGGGACACATAGCCGGACCGTTTGTCACTTACCGTGTGAAACCCCTAC  
 GACGACGGCATCGGCTGGACGAGTTGCTGACTGGCTCAACTCCCCAAGTGCAGGCTCCGGTTG  
 TCCAGCGGATGCCGACTACGGCGAGTGGCTGCGAGCGGTTGAGACTTCGCTGCCGTGCCGGATCG  
 CCAGCGCCACGCCCTGCTGCCCTGCTGCCGACAAACTACCGAGAGGCCGAAAGCGATATGCCGGTCA  
 ATCGCGCCCACCGACCGAGTTCCGCGCTGCCGTCCAAGCGAAAATCGGTCCGGACAAAGACATTCCGC  
 ACCTCACGGCGGACATCGCGAACATCAGCACCGTACTGCTGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:266)**

>gi|218754327|ref|ZP\_03533123.1| fatty-acid-CoA ligase fadD9  
 [Mycobacterium tuberculosis GM 1503]

MVELLPRFETITYRELWARAGTLATLASEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAP  
 PVTGLRPIVTEPTMIATSIDNLGDAVEVLAGHAPARLVDYHGKVDTREAVEAARARLAGSVTIDT  
 LAELIERGRALPATPIADSADDALALLIYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPGYPSTLNF  
 MPMSHVGGRQVLYGTLNSNGGTAYFVAKSDLSTLFEDLALVRPELCFVPRIDIWDMVFAEFHSEVDRRLVDG

**FIG. 8QQQ**

ADRAALEAQVKAELRENVLGGRFVMALTGSAPISAEMTAWVESLLADVHLVEGYGSTEAGMVLDGMVRR  
PAVIDYKLVDVPELGYFGTDQPYPRGEELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIAMAKVGPDQF  
VYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIIFIYGNNSARAYPLAVVPSGDAISRHGIENLKPVI  
SESLQEVARAAGLQSYEIPRDFIIETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADSQSNEL  
RELRLQSGPDAPVLPTLCRAAAALLGSTAADVRPDAHFADLGGDSLSALSLANLLHEIFGVDPVGIVSP  
ASDLRALADHIEAARTGVRRPSFASTHGRSATEVHASDLTLDKFIDAATLAAAPNLPAPSAQVRTVLLTG  
ATGFLGRYLALEWLDRMDLVNGKLICLVRARSDEEAQARLDATFDSDGPYLVRYRELGAGRLEVLAGDK  
GEADLGLDRVTVQRLADTVLIVDPAALVNHLVPLYSQFLFGPNAAGTAELLRLALTGKRKPYIYTSTIAVG  
EQIPPEAFTEDADIRAIAPTTRRIDDSYANGYANSKWAGEVLLREAHEQCGLPVTFRCDMILADTSYTGQ  
LNLPDMFTRLMLSIAATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYHVMNPY  
DDGIGLDEFVDWLNSPTSGSGCTIQRIADYGEWLQRFETSLRALPDRQRHASLLPLLHNYPREPAKPICGS  
IAPTDQFRAAVQEAKIGPDKDIPHITAIIAKYISNLRILLGLI

**ZP 03433592****Nucleotide sequence (SEQ ID NO:267)**

>gi|189214744:1-3126 Mycobacterium tuberculosis T85 NZ\_ABOW01000178,  
whole genome shotgun sequence

GACATCGCGCTGATCCGGTTGGCGCCGTGTCGGTTCCACTGCAGACCAGTGCGCCGGTCACCGGGTTGC  
GCCCGATCGTACCGAGACCGAGCCGACGATGATGCCACCAGCATGACAATCTGGCAGGCCGTGCA  
AGTGCCTGGCCGGTCACGCCCGCCGGCTGGTCGATTACGATTACACGGCAAGGTTGACACCCACCGC  
GAGGCCGTGCAAGCCGCCAGCTCGTTGGCCGGCTCGGTGACCATGACACACTGCCGAACGTGATCG  
AACCGCCAGGGCCTGCCGCCACACCCATTGCCGACAGCGCCGACGACGCGTGGCGTGTGATTAA  
CACCTCGGGTAGTACCGCGCACCCAAAGGCACATGTATCGCAGAGGCCAGGTGATGAGCTTCTGGCGC  
AAGTCGAGTGGCTGGTCAGGCCGAGCGGTTACCCCTCGATCACGCTGAACCTCATGCCGATGAGCCACG  
TCGGGGGCCGTCAGGTGCTCTACGGGACGCTTCAACGGCGTACCGCTACTTCGTCGCCAAGAGCGA  
CCTGTCACGCTGTTGAGGACCTGCCCTGGTGCAGGCCACAGAATTGTGCTTCGTGCCGCGCATCTGG  
GACATGGTGGTCGAGAGTCCACAGCGAGGTGACCGCCCTGGTGGACGGCGCGATCGAGCGGCC  
TGGAAAGCCAGGTGAAGGCCGAGCTGCCGGAGAACGTGCTGCCGGACGGTTGTCATGGCGCTGACCGG  
TTCCCGCCGATCTCCGCTGAGATGACGGCGTGGTCAGTCCTGCTGGCCGACGTGCATTGGTGGAG  
GGTACGGCTCCACCGAGGCCGGATGGTCTGAACGACGGCATGGTGCAGGCCGCCGCGGTGATCGACT  
ACAAGCTGGTCACGTGCCGAGCTGGCTACTCGGCACCGATCACCCCTACCCCCGGGGCAGCTGCT  
GGTCAAGACGCAAACCATGTTCCCCGGCTACTACCGAGCCGGATGTCACCGCCGAGGTGTTGACCCC  
GACGGCTTCTACCGGACGGGGACATCATGGCAAAGTAGGCCCGACCGAGTCGCTACCTCGACCGCC  
GCAACAAACGTGCTAAAGCTCTCCCAGGGCAGTCATGCCGTGTCGAAGCTCGAGGCCGTTGCG  
CAGCCCCTGGTCCGACAGATCTTCATCTACGGAACAGTCAGCCGGCTTACCGCTGGCGTGGTTGTC  
CCGTCGGGGACCGCCTTCGCCATGGCATCGAGAATCTCAAGCCCGATCGAGCTCCCTGCAGG  
AGGTAGCGAGGCCGGCCGGCTGCAATCCTACGAGATTCCACGCGACTTCATCATCGAAACACGCC  
CACCTGGAGAACGCCCTGCTCACCGCAGTCAGCTGGCACGCCAGTCAGGCCGAGGTCTATGGC  
GAACGTCTCGAGCGGCTCTACCGAGCTGGCGATAGCCAATCCAACGAGCTGCCGAGCTGCC  
GGCGTCCCGATGCCGGTGCCTCCGACCTGGTGGTGAACGCTCGCTCGGCCGCTGTTGGCT  
GGATGTGCGGCCGGACCGCAGTCAGCTGCCGACCTGGTGGTGAACGCTCGCTCGGCCGCTGTTGG  
CTGCTGCACGAGATCTCGCGCTGCGACGTGCCGGTGGTGTCAATTGTCAGCCGCCAGCGAC  
CCCTGGCCGACACATCGAACGAGCGCAGCCGGCTCAGGCCACCCAGCTTCGCCCTCGATACCGCT  
CTCCCGCAGCGAAGTGCACGCCAGCGACCTGCCGACCTGGTGGTGAACGCTCGCTCGGCC  
GCCCGAACCTGCCGGCACCGAGCGCCAAGTGCACGCCAGCGTACTGCTGACCGGCCACCG  
GTCGCTACCTGCCGCTGGAAATGGCTGACCGCATGGACCTGGTCAACGCAAGCTGATCTGC  
CGCCAGATCCGACGAGGAAGCACAAGCCGGCTGGACGCGACGTTGATAGCGGCCACCGT  
CGGCACCTACCGCGAATTGGCGCCGGCGCTCGAGGTGCTGCCGGCACAAGGGCGAGGCC  
GCCTGGACCGGGTACCTGGCAGCGGCTAGCCGACACGGTAGCTGTCAGGCCGCCCTGGT

**FIG. 8RRR**

CAACCACGTGCTGCCGTATGCCAGCTGTCGGCCAAACGCCGGCGGGACCGCCGAGTTGCTCGGCTG  
 GCGCTGACCGCAAGCGAAGCCATACATCTACACCTCGACGATCGCCGAGCAGATCCCGCCGG  
 AGGCCTCACCGAGGACGCCACATCCGGCCATCAGCCGACCCGAGGATCGACGACAGCTACGCCAA  
 CGGCTACCGCAACAGCAAGTGGGCCGGCAGGTGCTGCTGCCGAAGCTCACGAGCAGTGC  
 GTGACGGCTTCCGCTGCGACATGATCCTGGCCACACCAGCTATACCGGTAGCTCAACCTGCCGG  
 TGTCACCCGGCTGATGCTGAGCCTGCCGCTACCGGCATCGCACCCGGTCGTTCTATGAGCTGGATGC  
 GCACGGCAATCGGCAACGCCCACTATGACGGCTGCCGGTCGAATTGTCGAGAACGCATTGACC  
 CTTGGGACACATAGCCCAGCGTTGTCACCTACCAGTGATGAACCCCTACGACGGCATGGC  
 TGGACGAGTTCGTCGACTGGCTCAACTCCCCAAGCTAGCGGGTCCGGTTGCACGATCCAGCGG  
 CTACGGCGAGTGGCTGAGCGGTTGAGACTTCGCTGCCGCTTGCCGATGCCAGCGCCACGCC  
 CTGCTGCCCTGCTGACAACACTACCGAGAGCCTGCAAAGCGATATGCCGGTCATCGGCCACCGACC  
 AGTCCCGCCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCG  
 GATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGCTGCTGTGA

**Amino acid sequence (SEQ ID NO:268)**

>gi|215446840|ref|ZP\_03433592.1| fatty-acid-CoA ligase fadD9  
 [Mycobacterium tuberculosis T85]

DIALIRLGAVSVPLQTSAPVTGLRPIVTETEPTMIATSIDNLGDAVEVLAGHAPARLVVF DYHGKV DTHR  
 EAVEAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTGAPKGAMYRESQVMSFWR  
 KSSGWFEPSGPSITLNFMPPMSHVGRQVLYGTLNSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPR  
 DMVFAEFHSEVDRRLVGDADRAALEAQVKAELENVVLGGRFVMALTGSAPISEMATAWVESLLADV  
 HLYGSTEAGMVNLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDV  
 TAEVFD P DGFYRTGDIAMAKVGPDQFVYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN  
 SARA YPLAVVV PSGDALSRHGIENLKPVISES LQEVARAAGLQSYEIPRDFIETTPFTLENGLLTGIRKL  
 ARPQLKKFYGERLERLYTELADSQSNELRELRSQGPDAVLPTLCRAAAALGSTAADVRPD  
 AHFADLGGDSLSALSLAN LLHEIFGVDPVPGVIVSPASDLRALADHIEARTGVRRPSFASI  
 HGRSATEVHASDLTLDKFIDAATLAA APNLPAPSAQVRTVLLTGATGFLGRYL  
 ALEWLDRM DALVNKGKLTCLVRARSDEEAQARLDATFDSDGDPYLV RHYREL  
 GAGRLEVLAGDKGEADLGLDRVTWQR LADTVL D LIVDPAALVN  
 HVLPYSQ LFGPNAAGTAELLRL ALTGKRKP YIYTSTIAVGEQT  
 PPEAFTEDADIRAI SPTRRIDDSYANGYANSK  
 WAGEVLLREAHEQCGLP VTVFRCDMILADTSYTQQLNLPDMFTRLMLSLAATG  
 IAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICT LGTHSPDRFV  
 TYHVMNPYDDGIGLDEFVDWLNSPTSGSGCTI  
 QRIADYGEWLQRFETSLRALPDRQRHAS LLPLLHNYREPAK  
 PICGSIAPT DQFRAAVQEAKIGPDKDIPH  
 LTAAIIAKYISNLRLG LLL

**ZP\_03537669****Nucleotide sequence (SEQ ID NO:269)**

>gi|192384126:19547-22546 Mycobacterium tuberculosis T17  
 NZ\_ABQH01000288, whole genome shotgun sequence

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAGCGACGCCAGTTGCC  
 CCGCCAGTCCCAACGAGGCATCACCCAGGCATCGACCAGCCGGGTCGCGCTTCCACAGCTCATCC  
 TATGGTCATGGAGGGCTACGCCGATCGGCCGACTCGGCCAGCGTGCCTCCGCTCGTCACCGACCC  
 GACAGCGGCCGACCATGGTCGAGCTACTGCCGGTTCGAGACCATCACCTACCGGA  
 ACTGTGGGCC GCGCAGATGGCCACCGCGTTGAGCGCTGAGCCCGCGATCCGGCGGGCGACCGGTT  
 TGCCTGGCTGGCGCCGATCGTCACCGAGACCGAGCCGACGATGATCGCA  
 CCAGCATCGACAATCTGGCGACGCCGTCGAAGTGC  
 TGGCCGGTCACGCCGGCTGGCGTATT

**FIG. 8SSS**

CGATTACCACGGCAAGGTTGACACCCACCGCGAGGCCGTCGAAGCCGCCAGCTCGGTTGGCCGGCTCG  
 GTGACCATCGACACACTTGCAGAAGTGTGAGCTCGAGTGCTGGTCAAGCTGAGGCGCTGCCGCCACACCCATTGCCGACA  
 GCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACCCAAAGGCGCCATGTA  
 TCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGAGTGCTGGTCAAGCTGAGGCGAGCGGTTACCCCTCG  
 ATCACGCTGAACCTCATGCCATGAGCCACGTCGGGGCGCTCAGGTGCTCTACGGGACGCTTCCAACG  
 GCGTAGCCCTACTTCGTCGCCAAGAGCGACCTGTCACGCTGTCAGGAGCTCCACAGCGAGGTGACCGC  
 CACAGAATTGPGCTTCGTCGCCGCACTGGGACATGGTGTGCAAGAGTCCACAGCGAGGTGACCGC  
 CGCTTGGTGGACGGGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCAGCTGCCGGAGAACGTGC  
 TCGCGGAGCGTTGTCATGGCGCTGACCGGTTCCGCGCCGATCTCGCTGAGATGACGGCGTGGGTCGA  
 GTCCCTGCTGGCGACGTCGTCATTGGTGGAGGGTTACGGCTCCACCGAGGCCGGATGGTCTGAACGAC  
 GGCATGGTGGCGGCCCGCGTGTGACTACAGCTGGTCAAGCTGCGACGTCGGCTACTTCGGCA  
 CCGATCAGCCCTACCCCCGGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGCTACTACCAGCG  
 CCCGGATGTCACCGCCGAGGTGTTGACCCGACGGCTTCAACCGGACCGGGGACATCATGGCAAAGTA  
 GGCCCCGACCAGTTGCTCACCTCGACCGCCGCAACACGTCGCTAAAGCTCTCCAGGGCGAGTCATCG  
 CCGTGTGCAAGCTCGAGGCCGTTGGCGACAGCCGCTGGTCCGACAGATCTTCATCACGGAACAG  
 TGCCCGGGCCTACCCGCTGGCGTGGTGTGGTCCGGTCCGGGACGCGCTTCTCGCCATGGCATGAGAAT  
 CTCAAGCCCGTGTGATCAGCGAGTCCTGCACTGAGGAGTAGCGAGGGCGCCGCGCTGCAATCCTACGAGATT  
 CACCGCAGCTCATCATCGAACACAGCCGTTCACCGCTGGAGAACGGCCTGCTCACGGCATCGAACAG  
 GGCACGCCCAGTTGAAGAAGTTCTATGGCAACGTCCTGAGCGGCTCTATACCGAGCTGGCGATAGC  
 CAATCCAACGAGCTGCGGAGCTGCGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGCGTG  
 CCCGGCTGCGTTGCTGGCTCACCGCTGCGGATGTCGGGGGACGCGCAGCTCGCCGACCTGGTGG  
 TGACTCGCTCOGGCGCTGCGTGGCAACCTGCTGCAAGAGATCTCGCGTGCACGTGCGGGTGG  
 GTCATTGTCAGCCCGCAAGCGACCTGCGGGGCGTGGCGACACATGAAAGCAGCCGACCGCGTCA  
 GGGGACCCAGCTCGCTCGATAACACGGTCGCTCCGCGACCGAAGTGCACGCCAGGACCTCACGCTGGA  
 CAAGTTCATCGACGCTGCCACCGTGGCGCAGCCCCGAACCTGCCGGCACCGAGCGCCCAAGTGCAC  
 GTACTGCTGACCGGGCCACCGCTTTGGGTCGCTACCTGGGGCTGGAATGGCTGACCGCATGGACC  
 TGGTCAACGGCAAGCTGATCTGCCCTGGTCCGCGCAGATCCGACGAGGAAGCACAAGCCGGCTGGACGC  
 GACGTTGATAGCGGCGACCGTATTGGTGGGCACTACCGCGAATTGGGCGCCGCCCTGAGGTG  
 CTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCTGGACCGGGTCAACCTGGCAGCGCTAGCCGACACGG  
 TGGACCTGATCGTGGACCCCGCGGCCCTGGTCAACCGACGTCGCTGCCGTATAGCCAGCTGTTGGCCCAA  
 CGCCGGGGCACCGCGAGTTGCTTGGCTGGCGTGCACCGCAAGCGCAAGCCATACATCACACCTCG  
 ACGATCGCGTGGGCGAGCAGATCCGCCGGAGGCCTTACCGAGGACGCCGACATCCGGGCCATCAGCC  
 CGACCCCGCAGGATCGACGACAGTACGCCAACGGCTACCGCAACAGCAAGTGGGCGGGAGGTGCTGCT  
 GCGCGAAGCTCACGAGCAGTGCDDCTGCGCTGACGGTCTCCGCTGCCGACATGATCTGGGCCACAC  
 CAGCTATACCGGTAGCTCAACTGCCGGACATGTCACCCGGCTGATGCTGAGCCTGGC

**Amino acid sequence (SEQ ID NO:270)**

>gi|219558593|ref|ZP\_03537669.1| fatty-acid-CoA ligase fadD9  
 [Mycobacterium tuberculosis T17]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQADQPGVALPQLIRVMMEGYADRPALGQRALRFVTDP  
 DSGRTMELLPRFETITYRELWARAGTLATALESAEPAIRPGDRVCVLGFNSVDYTTID1ALTRLGAVSP  
 LQTSPVTLRPIVTETEPTMIATSDINLDAVEVLAGHAPARLVFDYHGKVDTHREAVEAARARLAGS  
 VTIDTIAELTERGRALPATPTADSADIALITYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPSGYPS  
 ITLMFMPMSHVGGRQVLYGTLNSNGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR  
 RLVDGADRAALEAQVKAELENVVLGGRFVMAUTGSAPIESAEMTAWVESLLADVHLVEYGSTEAGMVLD  
 GMVRRPAVIDYKLVDVPELGYFGTDQPYPRCELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKV  
 GPDQFVYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIFYGNSARAYPLAVVVPSGDALSRRHGIEN  
 J, KPVISESTI, QFVARAAGL, QSYET, PRDFIIETTPFTLENLITGIRKLARPQLKKFYGERLERLYTELADS  
 QSNELRELRQSGPDAPVLPTELCRAAAALLGSTAADVRPAHFADLGGDSLSALSLANLLHEIFGVDPVG  
 VIVSPASDLRALADHIEARTGVRRPSFASTHGRSATEVHASDLTLDKFIDAATLAAAPNLAPSQVRT  
 VLLTGATGFLGRYLALEWLDRMDLVNGKLICLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEV

**FIG. 8TTT**

LAGDKGEADLGLDRVTWQRLADTVDLIVDPAALVNHLPYSQLFGPNAAGTAELLRLALTGKRKPYIYTS  
TIAVGEQIPPEAFTEDADIRAIISPTRRIDDSYANGYANSKWAGEVLLREAHEQCGLPVTFRCDMILGRH  
QLYRSAQPAGHVTRADAEPG

**ZP 05224908**

**Nucleotide sequence (SEQ ID NO:271)**

>gi|163719654:2489-6013 Mycobacterium intracellulare ATCC 13950  
NZ\_ABIN01000072, whole genome shotgun sequence

ATGTCGACTGCCATTATGACGAACACCTCGACCGTCGCATCGAGGAACATCGCCAACGACCCCCAAT  
TCGCCGCCGCCCCGACCGGACCCGGCATCACCGCCGCCACCGAACCGCCGGGGCTGCGGCTGCCGCAGAT  
CATCCGGACCGTGTGCTGACGGCTACGCCGACCGGCCCTGCCCTGGCGCAGCGCGTCGTGGAGTTCTGTCACC  
GACGCCAAGACCGGGCGGACGACGCCGAGCTCTCCCCGTTTGAGACCATCACGTATGGCGAACTCG  
GCGAACGGGTTTCGGCCCTGGCCGTGCGCTGGGGCGCGACCGGGTGCACGGGTGCGCCCCGGGACCGCGCTCG  
GCTCGGCTTCAACAGCGTTGACTACGCCACCATCGACATCGCGCTGGGACCATCGGGCCGTGTCGGTG  
CCGCTGAGCAGACCGCGCGATCTCTCGTTGAGCCGATCGTCGCCAGACCGAGCCAGCCTGATCG  
CCTCGAGCGTCAACCAGCTGCCGACCGGGTGGAGCTGATCTGGCCGGGACCGACGTGCCCGCAAGCT  
CGTCGTGTCGACTACCAGCCCCAGTCGACGACCGACGCCGAGGGCTGGACCCCGCCGCCGTTG  
GCCGACTCCGGCGTCGGTGCAGGCTCTCGCCGACGTGCTGCCGCGGGCAAGGACCTGCCGGCGTC  
AGCCGCCGGCGAGCGACGAGACTCGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGCGCGCCAA  
GGGCGCGATGTACCCGAGAGCAACCGTGGCAAGATGTGGCCGGCGGGAGCAAGAACGACTGGTTGGGGAA  
AGCCGCCGCGTCGATCACCTCAACTTATGCGGATGAGCCACGTGATGGGGCGCGGAATCTCTACGGCA  
CGCTGGCGAACGGCGGCCACCGCGTACTTCGCCGCCCGCAGCGACCTGCTCCACCGCTCGAGGACCTCGA  
GTTGGTGC GGCCCACCGAGATGAACCTCGTCCCCCGCATCTGGGAGACGCTGTACGGCGAATTCCAGCGC  
CAGTCGAGCGCGGCTGGCGACGGCGATGCCGGAGGGCCCGAGACTGTGGCGCTGCCGT  
TGGAAAGAACAGGCCAGTACCTGCTGGCGGGCGGTTATCTCGCGATGACGGCTCGGCACCCACCTC  
GCCGGAGCTCAAGCGTGGCGAGTCGCTCTCGCAGATGACCTGATGGACGGCTACGGCTCCACCGAG  
GCCGGAATGGTGGTGTGTCGACGGGAGATTAGCTGGCGCCGTTATTGATTACAAGCTGGTGCACGTT  
CGGATCTGGCTATTCAGCACCGACCGTCCGCATCCGCGCGGTGAGTTGCTGCCGACCGAGAACAT  
GTTCCGGGTTATTACAAGCGGGCGAGACCACCGCGAACGTTGCTGACGAGGGTATTACCGCACC  
GGTGACGTGTTGCGCAGATCGCGCCGACCGGGCTGGTGTATGTCGATGCCGAAACAACGTGCTCAAGT  
TGGCCCAGGGCGAGTCGTCGACGCTGCCAAGCTGGAGGGCGTGTGCCAACAGCCGCTGATCCGCCA  
GATCTACGTTACGCCAACAGCTCCAGCCCTACCTGCTGCCGTGGTGGCGACCGAGGAAGCGTTG  
GCCGACAACGATCTGAGTCGCTCAAGCCGAAGATGCCGACTCGCTGAGAACGGTCCGCCAAGGAGACCG  
GCCCGCAGTCCTACGAGGTGCCGCCGCGACTTCATCATCGAGACCGCCGTTACCCCTGGAAAACGGCCT  
GCTGACCGGGATCCGCAAGCTGGCGTGGCCAAGCTCAAGGGCGACTACGGGGATCGGCTCGAGCAGATG  
TATGCCGAGCTGCCGCCGGACAGGCCAACAGAGTTGGCGAATGCCGCCGAGCGGGCGGCCGG  
TGGCCCAGACCGTGAGCGGGCGCGCCCTGCTGGGTGCGACGGCGGGGATCTGCGCAGATGC  
CCACTTCACCGATCTGGTGGAGACTCGTGTGCGCGTGCAGCCGCCAACGACCTGGCGGGATGCCGCC  
GATGTCGACGTGCCGGTGGGGGTGATCGTCAGCCGCCAACGACCTGGCGGGATGCCGCC  
AGGCCGAGCGGCCAGGGCTCAAGGCCGCCGACGTTGCCGCCGTGACGGTGCACGGTGCACGGTGC  
CGCCAGCTGACCTCACGCTGACAAGTCTCGACGAGCGACCCCTGCCGCCGCCAGCCTGCCAAG  
CCGGCCACCGAGGTGCGCACCGTGTGACCGGGCGACCGGCTTTTGGCCGCTACCTGGCGCTGG  
ACTGGCTCGAGCGGGATGGACATGGTCGACGGCAAGGTGATGCCCTGGTGCACGGGCCACCGATGAGGA  
GGCGCCGCCGGCTGGACAAGACCTCGACAGCGGCCACCCAAAATGCTGGCGCAACTACCGGGCTG  
GCCGCCGACCCACCTCGAGGTGATGCCGCCGACAAGGGTGAAGGCAACCTCGGCCCTGGACCCCCAGACCT  
GGCAGCGACTGGCGAGGGAGGTGACGCTGATCGTCACCCGCCGCCGCTGGTCAACCACGTGCTGCC  
CAGCGAGCTGTTGCCGCCAACGCCCTGGGCACCGCGGAGCTGATCCGATCGCGCTGACCTCAGGCAA  
AAGCCCTACACCTACGTGTCGACGATGGGGTGGCGATCAGATCCGAGGTGAGTTGCTCGAGAACG

**FIG. 8UU**

CCGACATCCGCCAGATCAGGCCACCCCGAGATCAACGACGGCTACGCCAACGGTACGGCAACAGCAA  
GTGGGCCGGCGAGGTGTTGCTGCGCGAGGCCACGACCTGTGCGGCCTGCCGTACGGTGTCCGCTGC  
GACATGATCCTGGCCGACACCACCTATGCCGGCAGCTCAACCTGCCGACATGTTCACCCGGCTGATGC  
TGAGCCTGGTCGCCACCGGTATCGCCCCGGTCGTTCTACGAACGGACGCCACGGCAACCGCCAGCG  
GGCACACTACGACGGTTGCCGGTCGAGTTCATCGCCGGCGATCTCGACGCTGGGACCCAAATCACC  
GACAGCGACACGGGCTTCAGACCTACCGTGTGATGAAACCCCTACGACGACGGCACTGGGCTGGATGAGT  
ACATCGATTGGCTGATCGAGGCCGGTATTGATCGAGCGGATGCCGATTACTCCGAGTGGCTGCCGCG  
CTTCGAGACCTCGCTGCCGCGCTGCCGATGCCGAGCGTCAGTACTCGCTGCCGCTGCTGCACAAC  
TACCGAGACGGGAAAGCCGATCAACGGCTCGATGCCGCCACCGACGTGTCGGTGCAGG  
AAGCGAAAATCGGCCCGACAAAGACATCCCGACGTCTGCCGCCGTGATCGTAAGTACATCACCAG  
CCTGGAGTTGCTCGGACTCCTCTGA

**Amino acid sequence (SEQ ID NO:272)**

>gi|254819907|ref|ZP\_05224908.1| FadD9 [Mycobacterium intracellulare ATCC 13950]

MSTAIHDEHLDRIEELIANDPQFAAARPDPAITAATEAPGLRLPQIIRTVLDGYADRPALAQRVVEFVT  
DAKTGRRTAELLPRFETITYGELGERVSALGRAWAGDAVRPGDRVCVLGFNSVDYATIDIALGTIGAVSV  
PLQTSAISSLQPIVAETEPSLIASSVNQLPDAVELILAGDHVPKGKLVFDYQPQVDDQREAVEAAAARL  
ADSGVAEALADVLRRGKDLPAVEPPASDEDSLALLIYTGSTGAPKGAMYPQSNGKMWRRGSKNWFGE  
SAASITLNFMPSHVMGRGILYGTGNGGTAYFAARSDLSTILLEDLELRPTEMNFVPRIWETLYGEFQR  
QVERRLADGDAGPEARETVAAAVLEEQRQYLLGGRFIFAMTGSAPTSPELKAWAESLLQMHLMDGYGSTE  
AGMVLFDGEIQRPPVIDYKLVDVPLGYFSTDPRPHPRGELLRTENMFPGYYKRAETTANVFDEDGYYRT  
GDVFAETIAPDRLVYVDRNNVLKLAQGEFVTIYLAKLEAVFGNSPLIRQIYVGNSSQPYLLAVVPTEEAL  
ADNDLESLKPKIADSLQKVAKETGLQSIEVPRDFIETTPFTLENLGIRKLAWPKLKAHYGDRLEQM  
YAELAAQANELAELRRSGAAAPVAQTVSRAAAALLGATAGDLSADAHTDLGGDSLSALTFGNLLRETF  
DVDVPVGIVVSPANDLAGITIAYIEAEERQGSKRPTFAAVHGRGATMVHASDLTLDFLDEATLAAAPSLPK  
PATEVRTVLLTGATGFLGRYLALDWLERMDMVDGKVIALVRARTDEEARARLDKTFDSDGDPKLLAHYQRL  
AADHLEVIAGDKGEANGLGLDPQTWQRLAEEVDVIVDPAALVNHVLPYSSELFGPNALGTAELIRIALTSRQ  
KPYTYVSTIGVDQIOPGEFVENADIRQISATREINDGYANGYGN SKWAGEVLLREAHDLCGLPVTVFRC  
DMILADTTYAGQLNLPDMFTRLMLSIVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAISTLGTQIT  
DSDTGFBTYHVMNPYDDGIGLDEYIDWLIEAGYSTERIADYSEWLRRFETSLRALPDRQRQYSLLPLLHN  
YQKPEKPIGSMAPTDVFRAAVQEAKIGPDKDIPHVSAPVIVKYITDLELLGLL

FIG. 9A

Accession Numbers are from NCBI; GenBank, Release 159.0 as of March 2008  
 EC Numbers are from KEGG, Release 42.0 as of April 2007 (plus daily updates up to March 2008)

CATEGORY	GENE	NAME	ACCESSION	EC NUMBER	MODIFICATION	USE	ORGANISM
1. Fatty Acid Production Increase / Product Production Increase							
<i>increase acyl-CoA</i>							
<i>reduce catabolism of derivatives and intermediates</i>							
<i>reduce feedback inhibition</i>							
<i>attenuate other pathways that consume fatty acids</i>							
		Acetyl-CoA carboxylase, subunit A (carboxyltransferase alpha)	AAC73296 NP_414727	6.4.1.2	Over-express	increase Malonyl-CoA production	<i>Escherichia coli, Lactococci</i>
	accA	Acetyl-CoA carboxylase, subunit B (BCCP; biotin carboxyl carrier protein)	NP_417721	6.4.1.2	Over-express	increase Malonyl-CoA production	<i>Escherichia coli, Lactococci</i>
	accB	Acetyl-CoA carboxylase, subunit C (biotin carboxylase)	NP_417722	6.4.1.2, 6.3.4.14	Over-express	increase Malonyl-CoA production	<i>Escherichia coli, Lactococci</i>
	accC	Acetyl-CoA carboxylase, subunit D (carboxyltransferase beta)	NP_416819	6.4.1.2	Over-express	increase Malonyl-CoA production	<i>Escherichia coli, Lactococci</i>
	accD	pyruvate dehydrogenase, subunit E1	NP_414656, AAC73226	1.2.4.1	Over-express	increase Acetyl-CoA production	<i>Escherichia coli</i>
	aceE						

FIG. 9B

aceF	pyruvate dehydrogenase, subunit E2	NP_ 414657 AAC75356, NP_ 416799	2.3.1.12 Over-express		increase Acetyl-CoA production	<i>Escherichia coli</i>
ackA	acetate kinase		2.7.2.1 Delete or reduce		increase Acetyl-CoA production	<i>Escherichia coli</i>
ackB	acetate kinase AckB	BAB81430	2.7.2.1 Delete or reduce		increase Acetyl-CoA production	<i>Escherichia coli</i>
acpP	acyl carrier protein	AAC74178	NONE	Over-express	increase Acetyl-CoA production	<i>Escherichia coli</i>
fadD	acyl-CoA synthase	AP_ 002424	2.3.1.86, 6.2.1.3	Over-express	increase Fatty acid production	<i>Escherichia coli</i> W3110
adhE	alcohol dehydrogenase	CAA47743	1.1.1.1, 1.2.1.10	Delete or reduce	increase Acetyl-CoA production	<i>Escherichia coli</i> W3111
cer1	Aldehyde decarbonylase	BAA11024	4.1.99.5	Over-express	increase Acetyl-CoA production	<i>Arabidopsis thaliana</i>
fabA	beta-hydroxydecanoyl thioester dehydrase	NP_ 415474	4.2.1.60	express	fatty acyl-CoA production	<i>E. coli</i> K12
fabD	[acyl-carrier-protein] S-malonyltransferase	AAC74176	2.3.1.39	Over-express	increase Acetyl-CoA production	<i>E. coli</i> K12
fabF	3-oxoacyl-[acyl-carrier-protein] synthase II	AAC74179	2.3.1.179	Delete or OverExpress	increase Acetyl-CoA production	<i>E. coli</i> K12
fabG	3-oxoacyl-[acyl-carrier-protein] reductase	AAC74177	1.1.1.100	Over-express	increase Acetyl-CoA production	<i>E. coli</i> K12
fabH	3-oxoacyl-[acyl-carrier-protein] synthase III	AAC74175	2.3.1.180	Over-express	increase Acetyl-CoA production	<i>E. coli</i> K12, <i>lactococci</i>

FIG. 9C

<i>fabI</i>	enoyl-[acyl-carrier-protein] reductase, NADH-dependent	NP_415804	1.3.1.9	express	fatty acyl-CoA production	<i>E. coli</i> K12, <i>lactococci</i>
<i>fabR</i>	Transcriptional Repressor (3R)-hydroxymyristol acyl carrier protein dehydratase	NP_418398	NONE	Delete or reduce	modulate unsaturated fatty acid production	<i>E. coli</i> K12
<i>fabZ</i>	acyl-CoA dehydrogenase	NP_414722	4.2.1.-			<i>E. coli</i> K12
<i>fadE</i>	Fatty Acyl-CoA reductase	AAC73325	1.3.99.3, 1.3.99.-	Delete or reduce	Increase Acetyl-CoA production	<i>Acinetobacter</i> sp., i.e. <i>calcoaceticus</i>
<i>acr1</i>	Glutathione synthase biosynthetic sn-glycerol 3-phosphate dehydrogenase	YP_047869, AAC45217	1.2.1.42	Over-express	for fatty alcohol production	<i>E. coli</i> K12
<i>GST.gshB</i>	lactate dehydrogenase	P04425	6.3.2.3	Delete or reduce	Increase Acyl-CoA	<i>Saccharomyces cerevisiae</i>
<i>gpsA</i>	Triglyceride Lipase	AAC76632, NP_418065	EC: 1.1.1.94	Delete or reduce	Increase Acetyl-CoA production	<i>E. coli</i> K12
<i>ldhA</i>	Malonyl-CoA decarboxylase	AAC74462, NP_415898	EC: 1.1.1.27, 1.1.1.28	Delete or reduce	Increase Fatty acid production	<i>Saccharomyces cerevisiae</i>
Lipase	aspartate 1-decarboxylase	CAA89087, CAA98876	3.1.1.3	express		<i>Saccharopolyspora erythraea</i>
<i>panD</i>	pantothenate kinase	BAB96708	4.1.1.11	Over-express	Increase Acyl-CoA	<i>Escherichia coli</i> W3110
<i>panK</i> a.k.a. <i>coaA</i>		AAC76952	2.7.1.33	Over-express, Over-express, R106K mutation	Increase Acetyl-CoA production	<i>E. coli</i>
<i>panK</i> a.k.a. <i>coaA</i> , R106K		AAC76952	2.7.1.33		Increase Acetyl-CoA production	<i>E. coli</i>

EIG. 9

pdh	Pyruvate dehydrogenase formate acetyltransferase (pyruvate formate lyase)	BAB34380, AAC73226, NP_415392	1.2.4.1	Over-express	increase Acetyl-CoA production
pflB		AAC73989, P09373	EC: 2.3.1.54	Delete or reduce	increase Acetyl-CoA production
plsB	acyltransferase	AAC77011	2.3.1.15	D311E mutation	reduce limits on Acyl-CoA pool
poxB	pyruvate oxidase	AAC73958, NP_415392	1.2.2.2	Delete or reduce	increase Acetyl-CoA production
pta	phosphotransacetylase	AAC75357, NP_416800	2.3.1.8	Delete or reduce	increase Acetyl-CoA production
udhA	pyridine nucleotide transhydrogenase	CAA46822	1.6.1.1	Over-express	conversion NADH to NADPH or vice versa
fadB	fused 3-hydroxybutyryl-CoA epimerase/delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase/enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase	AP_003956	1.1.1.35	Delete or reduce	Block fatty acid degradation
fadJ	3-hydroxyacyl-CoA dehydrogenase; K01692 enoyl-CoA hydratase; K01782 3-hydroxybutyryl-CoA epimerase	AACT5401	1.1.1.35, 4.2.1.17, 5.1.2.3, 5.3.3.8,		Block fatty acid degradation
fadA	3-ketoacyl-CoA thiolase	BAE77458	2.3.1.16	Delete or reduce	Block fatty acid degradation
fadI	beta-ketoacyl-CoA thiolase	AAC75402	2.3.1.16	Delete or reduce	Block fatty acid degradation
YdiO	acyl-coA dehydrogenase	YP_852786	1.3.99,-	Delete or reduce	Block fatty acid degradation

FIG. 9E

2. Structure Control							
<b>2A. Chain Length Control</b>							
2	tesA	thioesterase	P0ADA1	3.1.2..1. 3.1.1.5	Delete and/or express	C18 Chain Length	
	tesA without leader sequence	thioesterase	AAC73596, NP_415027	3.1.2..1. 3.1.1.5	express or overexpress	C18:1	<i>E.coli</i>
	tesA without leader sequence:L109P	thioesterase	P0ADA1	3.1.2..1. 3.1.1.5	Express and/or overexpress mutation L109P	<C18 Chain Length	<i>E. coli</i>
	fatB1 (umbellularia)	thioesterase	Q41635	3.1.2.14	express or overexpress	C12:0	<i>Umbellularia californica</i>
	fatB2 (umbellularia)DE LET E umbelluria)	thioesterase	AAC49269	3.1.2.14	express or overexpress	C8:0 - C10:0	<i>Cuphea hookeriana</i>
	fatB3	thioesterase	AAC72881	3.1.2.14	express or overexpress	C14:0 - C16:0	<i>Cuphea hookeriana</i>
	fatB (cinnamomum)	thioesterase	Q39473	3.1.2.14	express or overexpress	C14:0	<i>Cinnamomum camphora</i>
	fatB[M141T]*	thioesterase	CAA85388	3.1.2.14	express or overexpress	C16:1	<i>Arabidopsis thaliana</i>
	fatA1 (Helianthus)	thioesterase	AAL79361	3.1.2.14	express or overexpress	C18:1	<i>Helianthus annuus</i>
	atfata (ARABIDOPSIS FAT A ACYL- ACP THIOESTERAS E)						
			NP_189147, NP_193041	3.1.2.14	express or overexpress	C18:1	<i>Arabidopsis thaliana</i>
	fatA	thioesterase	CAC39106	3.1.2.14	express or overexpress	C18:1	<i>Brassica juncea</i>

FIG. 9F

	fatA (cuphea) thioesterase	AAC72883	3.1.2.14	express or overexpress	C18:1	Cuphea hookeriana
2B. Branching Control	attenuate FabH					
	express FabH from <i>S. glaucescens</i> or <i>S. coelicolor</i> and knock out endogenous FabH				increase branched chain fatty acid derivatives	
	express FabH from <i>B. subtilis</i> and knock out endogenous FabH					
	<i>bdk</i> - <i>E3</i> - dihydrolipoyl dehydrogenase subunit			EC 1.2.4.4		
	<i>bkd</i> - <i>E1</i> -alpha/beta subunit	decarboxylase subunits of branched-chain $\alpha$ -keto acid dehydrogenase complex		EC 1.2.4.4		
	<i>bkd</i> - <i>E2</i> - dihydrolipoyl transacylase subunit			EC 1.2.4.4		
	<i>bkdA1</i>	branched-chain $\alpha$ -keto acid dehydrogenase $\alpha$ -subunit (E1a)	NP_628006	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors <i>Streptomyces coelicolor</i>

FIG. 9G

bkdB1	branched-chain $\alpha$ -keto acid dehydrogenase $\alpha$ -subunit (E1b)	NP_628005	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces coelicolor</i>
bkdC1	dihydriopyl transacetylase (E2)	NP_628004	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces coelicolor</i>
bkdA2	branched-chain $\alpha$ -ketoadic dehydrogenase $\alpha$ -subunit (E1a)	NP_733618	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces coelicolor</i>
bkdB2	branched-chain $\alpha$ -ketoadic dehydrogenase $\beta$ -subunit (E1b)	NP_628019	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces coelicolor</i>
bkdC2	dihydriopyl transacetylase (E2)	NP_628018	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces coelicolor</i>
bkdA	branched-chain $\alpha$ -ketoadic dehydrogenase $\alpha$ -subunit (E1a)	BAC72074	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>
bkdB	branched-chain $\alpha$ -ketoadic dehydrogenase $\beta$ -subunit (E1b)	BAC72075	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>
bkdC	dihydriopyl transacetylase (E2)	BAC72076	EC 2.3.1.168	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>
bkdF	branched-chain $\alpha$ -ketoadic dehydrogenase $\alpha$ -subunit (E1a)	BAC72088	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>
bkdG	branched-chain $\alpha$ -ketoadic dehydrogenase $\beta$ -subunit (E1b)	BAC72089	EC 1.2.4.4	express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>

FIG. 9H

	blkdh	dihydroliptoyl transacetylase (E2) branched-chain $\alpha$ -ketooacid dehydrogenase $\alpha$ -subunit (E1a)	BAC72090 NP 390285	EC 2.3.1.168 EC 1.2.4.4	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Streptomyces avermitilis</i>
	blkdaA	branched-chain $\alpha$ -ketooacid dehydrogenase $\alpha$ -subunit (E1a)	NP 390284	EC 1.2.4.4	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Bacillus subtilis</i>
	blkdAB	dihydroliptoyl transacetylase (E2) branched-chain $\alpha$ -ketooacid dehydrogenase $\alpha$ -subunit (E1b)	NP 390283	EC 2.3.1.168	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Bacillus subtilis</i>
	blkdbB	dihydroliptoyl transacetylase (E2) branched-chain $\alpha$ -ketooacid dehydrogenase $\alpha$ -subunit (E1a)	AAA65614	EC 1.2.4.4	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Pseudomonas putida</i>
	blkda1	branched-chain $\alpha$ -ketooacid dehydrogenase $\alpha$ -subunit (E1a)	AAA65615	EC 2.3.1.168	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Pseudomonas putida</i>
	blkda2	dihydroliptoyl transacetylase (E2)	AAA65617	1.8.1.4	express or Over-Express express or Over-Express	make branched-chain acyl-CoA precursors	<i>Escherichia coli</i>
	blkdc	dihydroliptoyl transacetylase (E2)	NP 414658	2.6.1.42	express or Over-Express express or Over-Express	make branched- $\alpha$ -ketooacids	<i>Escherichia coli</i>
lpd	dihydroliptoylamine dehydrogenase (E3)	branched-chain amino acid aminotransferase	YP 026247	2.6.1.42	express or Over-Express express or Over-Express	make branched- $\alpha$ -ketooacids	<i>Lactococcus lactis</i>
IlvE		branched-chain amino acid aminotransferase	AAF34406	2.6.1.42	express or Over-Express express or Over-Express	make branched- $\alpha$ -ketooacids	<i>Pseudomonas putida</i>
IlvE		branched-chain amino acid aminotransferase	NP 745648	2.6.1.42	express or Over-Express express or Over-Express	make branched- $\alpha$ -ketooacids	<i>Streptomyces coelicolor</i>
IlvE		branched-chain amino acid aminotransferase	NP 629657	2.6.1.42	express or Over-Express express or Over-Express	make branched- $\alpha$ -ketooacids	

FIG. 9I

ccr	crotomyl-CoA reductase	NP_630556	1.6.5.5.1. 1.1.1	express or Over- Express	Converting crotonyl-CoA to butyryl-CoA	<i>Streptomyces coelicolor</i>
ccr	crotomyl-CoA reductase	AAD53915	1.6.5.5.1. 1.1.1	express or Over- Express	Converting crotonyl-CoA to butyryl-CoA	<i>Streptomyces cinnamoneensis</i>
IcmA, isobutyryl- CoA mutase	isobutyryl-CoA mutase, subunit A	NP_629554	5.4.99.2	express or Over- Express	converting butyryl-CoA to isobutyryl-CoA	<i>Streptomyces coelicolor</i>
IcmA, isobutyryl- CoA mutase	isobutyryl-CoA mutase, subunit A	AAC08713	5.4.99.2	express or Over- Express	converting butyryl-CoA to isobutyryl-CoA	<i>Streptomyces cinnamoneensis</i>
IcmB, isobutyryl- CoA mutase	isobutyryl-CoA mutase, subunit B	NP_630904	5.4.99.2	express or Over- Express	converting butyryl-CoA to isobutyryl-CoA	<i>Streptomyces coelicolor</i>
IcmB, isobutyryl- CoA mutase	isobutyryl-CoA mutase, subunit B	CAB59633	5.4.99.2	express or Over- Express	converting butyryl-CoA to isobutyryl-CoA	<i>Streptomyces cinnamoneensis</i>
FabH, ACPS and fabF genes with specificity for branched chain acyl-CoAs						
IlvE	branched-chain amino acid aminotransferase	CAC12788	EC2.6.1.4 2	over express	branched chain amino acid aminotransferase	<i>Staphylococcus carnosus</i>
FabH1	beta-ketoacyl-ACP synthase III	NP_626634	2.3.1.180	express or Over- Express	initiation of branched-chain fatty acid biosynthesis	<i>Streptomyces coelicolor</i>

FIG. 9J

ACP	acyl-carrier protein	NP_626635	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	<i>Streptomyces coelicolor</i>	
FabF	beta-ketoacyl-ACP synthase II	NP_626636	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	<i>Streptomyces coelicolor</i>	
FabH3	beta-ketoacyl-ACP synthase III	NP_823466	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	<i>Streptomyces avermitilis</i>	
FabC3 (ACP)	acyl-carrier protein	NP_823467	NONE	express or Over-Express	initiation and elongation of branched-chain fatty acid biosynthesis	<i>Streptomyces avermitilis</i>	
FabF	beta-ketoacyl-ACP synthase II	NP_823468	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	<i>Streptomyces avermitilis</i>	
FabH A	beta-ketoacyl-ACP synthase III	NP_389015	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	<i>Bacillus subtilis</i>	
FabH B	beta-ketoacyl-ACP synthase III	NP_388898	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	<i>Bacillus subtilis</i>	
ACP	acyl-carrier protein	NP_389474	NONE	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	<i>Bacillus subtilis</i>	

FIG. 9K

		beta-ketoacyl-ACP synthase II	NP 389016	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	Bacillus subtilis
		beta-ketoacyl-ACP synthase III	ZP 0164305 9	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	<i>Stenotrophomonas maltophilia</i>
		acyl-carrier protein	ZP 0164306 3	NONE		initiation and elongation of branched-chain fatty acid biosynthesis	<i>Stenotrophomonas maltophilia</i>
	SmaIDRAFT_08 21	beta-ketoacyl-ACP synthase II	ZP 0164306 4	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	<i>Stenotrophomonas maltophilia</i>
	SmaIDRAFT_08 22	beta-ketoacyl-ACP synthase II	YP 123672	2.3.1.180	express or Over-Express	initiation of branched-chain fatty acid biosynthesis	<i>Legionella pneumophila</i>
	FabH	beta-ketoacyl-ACP synthase III	YP 123675	NONE		initiation and elongation of branched-chain fatty acid biosynthesis	<i>Legionella pneumophila</i>
	ACP	acyl-carrier protein				elongation of branched-chain fatty acid biosynthesis	<i>Legionella pneumophila</i>
	FabF	beta-ketoacyl-ACP synthase II	YP 123676	2.3.1.179	express or Over-Express	elongation of branched-chain fatty acid biosynthesis	<i>Legionella pneumophila</i>
	FabH	beta-ketoacyl-ACP synthase III	NP 415609	2.3.1.180	delete or reduce	initiation of branched-chain fatty acid biosynthesis	<i>Escherichia coli</i>

FIG. 9L

To Produce Cyclic Fatty Acids						
	FabF	beta-ketoacyl-ACP synthase II	NP_415613	2.3.1.179	delete or reduce	elongation of branched-chain fatty acid biosynthesis
AnsJ		dehydratase (putative)	not available	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
AnsK		CoA ligase (putative)	not available	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
AnsL		dehydrogenase (putative)	not available	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
ChcA		enoyl-CoA reductase	UT2144	EC 1.3.1.34	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
AnsM		oxidoreductase (putative)	not available	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
PlmJ		dehydratase (putative)	AAQ84158	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
PlmK		CoA ligase (putative)	AAQ84158	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
PlmL		dehydrogenase (putative)	AAQ84159	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
ChcA		enoyl-CoA reductase	AAQ84160	EC 1.3.1.34	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis
PlmM		oxidoreductase (putative)	AAQ84161	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis

FIG. 9M

	ChcB	enoyl-CoA isomerase	AF268489	not available	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis	<i>Streptomyces collinus</i>
	ChcB/CaiD	enoyl-CoA isomerase	NP_629292	4.2.1.-	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis	<i>Streptomyces coelicolor</i>
	ChcB/CaiD	enoyl-CoA isomerase	NP_824296	4.2.1.-	express or Over-Express	cyclohexylcarbo nyl-CoA biosynthesis	<i>Streptomyces avermitillii</i>
<b>2C. Saturation Level Control</b>							
	Sfa	Suppressor of FabA	AAN79592, AAC44390	NONE	Over-express	increase monounsaturated fatty acids	<i>E.coli</i>
	also see FabA in sec. 1				express	produce unsaturated fatty acids	
	GnsA	suppressors of the secG null mutation	ABD18647.1	NONE	Over-express	increase unsaturated fatty acid esters	<i>E.coli</i>
	GnsB	suppressors of the secG null mutation	AAC74076.1	NONE	Over-express	increase unsaturated fatty acid esters	<i>E.coli</i>
		also see section 2A - items with :0 are unsaturated (no double bonds) and with :1 are saturated ('1 double bond)					
	fabB	3-oxoacyl-[acyl]-carrier-protein synthase I	BAA16180	EC:2.3.1. 41	overexpress	modulate unsaturated fatty acid production	<i>Escherichia coli</i>

FIG. 9N

	fabK	trans-2-enoyl-ACP reductase II	AAF98273	1.3.1.9	express	modulate unsaturated fatty acid production	<i>Streptococcus pneumonia</i>
	fabL	enoyl-(acyl carrier protein) reductase	AAU39821	1.3.1.9	express	modulate unsaturated fatty acid production	<i>Bacillus licheniformis DSM 13</i>
	fabM	trans-2, cis-3-decenoyl-ACP isomerase	DAA05501	4.2.1.17	Over-express	modulate unsaturated fatty acid production	<i>Streptococcus mutans</i>
<u>Fatty Aldehyde Output</u>							
	thioesterase	see chain length control section			express	produce	
<u>Export</u>							
	Wax ester exporter (FATP family, Fatty Acid (long chain) Transport Protein)						<i>Drosophila melanogaster</i>
	ABC transport protein	putative alkane transporter	NP_524723	NONE	express	export wax	<i>Rhodococcus erythropolis</i>
	CER5	wax transporter	AAN73268	NONE	express	export products	<i>Arabidopsis thaliana</i>
	AtMRP5	Arabidopsis thaliana multidrug resistance-associated	Atg51500, AY734542, Atg21090, Atg51460	NONE	express	export products	<i>Arabidopsis thaliana</i>
	AmiS2	ABC transporter AmiS2	NP_171908	NONE	express	export products	<i>Rhodococcus sp.</i>

FIG. 9O

	AtPGP1	ARABIDOPSIS THALIANA P GLYCOPROTEIN1	NP_181228	NONE	express	export products	Arabidopsis <i>thaliana</i>
	AcrA	putative multidrug-efflux transport protein acrA	CAF23274	NONE	express	export products	<i>Candidatus</i> <i>Protochlamydia</i> <i>amoebophila</i> UWE25
	AcrB	probable multidrug- efflux transport protein, acrB	CAF23275	NONE	express	export products	<i>Candidatus</i> <i>Protochlamydia</i> <i>amoebophila</i> UWE25
	TolC	Outer membrane protein [Cell envelope biogenesis. transmembrane protein affects septum formation and cell membrane permeability	ABD59001	NONE	express	export products	<i>Francisella</i> <i>tularensis subsp.</i> <i>novicida</i>
	AcrE	Acriflavin resistance protein F	YP_312213	NONE	express	export products	<i>Shigella sonnei</i> Ss046
	AcrF		P24181	NONE	express	export products	<i>Escherichia coli</i>
	tlI1618	multidrug efflux transporter	NP_682408.1	NONE	express	export products	<i>Thermosynechoc</i> <i>occus elongatus</i> BP-11
	tlI1619	multidrug efflux transporter	NP_682409.1	NONE	express	export products	<i>Thermosynechoc</i> <i>occus elongatus</i> BP-11
	tlI0139	multidrug efflux transporter	NP_680930.1	NONE	express	export products	<i>Thermosynechoc</i> <i>occus elongatus</i> BP-11
<u>5. Fermentation</u>							increase output efficiency

FIG. 9P

umuD	DNA polymerase V, subunit	YP_310132	3.4.21.-	Over-express	increase output efficiency	<i>Shigella sonnei</i> Ss046
umuC	DNA polymerase V, subunit	ABC42261	2.7.7.7	Over-express	increase output efficiency	<i>Escherichia coli</i>
	NADH:NADPH transhydrogenase (alpha and beta subunits) (pntA, pntB)	P07001. <u>P0AB70</u>	1.6.1.2	express	increase output efficiency	<i>Shigella flexneri</i>

FIG. 10A

	Annotations	Pfam	Accession #
Zn-dep. Alcohol Dehydrogenases			
YjgB	best homolog of <i>Acinetobacter alrA</i> (50%)	00107, 08240	NP_418690
YahK	best homolog of yjgB (31%)	00107, 08240	NP_414859
AdhP	active on ethanol, more efficient as aldehyde reductase	00107, 08240	NP_415995
YdjL	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP_416290
YdjJ	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP_416288
YjgV (idnD)	oxidizes L-lidonate using NAD and reduces of 5-ketogluconate using NADH or NADPH	00107, 08240	NP_418588
Tdh	converts threonine and NAD to 1,2-amino-3-oxobutanate and NADH	00107, 08240	NP_418073
YjjN	possible L-galactonate oxidoreductase	00107, 08240	NP_418778
RspB	predicted oxidoreductase, Zn-dependent and NAD(P)-binding, unknown substrate	00107, 08240	NP_416097
GatD	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding	00107, 08240	NP_416594
YphC	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP_417040
YhdH	predicted oxidoreductase, Zn-dependent and NAD(P)-binding (3D)	00107, 08240	NP_417119
YcjQ	predicted oxidoreductase, Zn-dependent and NAD(P)-binding, large operon	00107	NP_415829
YncB	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107	NP_415966
Qor	NADPH:quinone reductase and related Zn-dependent oxidoreductases	00107, 08240	NP_418475
Adh3 (frmA)	alcohol dehydrogenase class III/glutathione-dependent formaldehyde dehydrogenase	00107, 08240	NP_414890
YbdR	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	NP_415141
YggP	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	00107, 08240	YP_026187
Fe-dep. Alcohol Dehydrogenases			
YiaY	unknown function, operon with aldB (aldehyde dehydrogenase)	00465	YP_026233
FucO	glyceraldehyde to ethylene glycole or lactaldehyde to 1,3-propanediol	00465	NP_417279
EutG	ethanolamine utilization?	00465	NP_416948
YqhD	reduces butanal, response mechanism to lipid peroxygenation? (3D)	00465	NP_417484
AdhE	three Fe++-dependent catalytic functions: alcohol dh, acetaldehyde dh and pyruvate formate-lyase deactivate, poorly active with pentanol	00465	NP_415757
Aldo-Keto Reductases			
yafB (dkgB)	reduces methylglyoxal, 2,5-diketogluconate and 4-nitrobenzaldehyde	00248	NP_414743
YdjG	reduces methy glyoxal, NADH specific	00248	NP_416285
YeaE	reduces methy glyoxal	00248	NP_416295
YqhE (dkgA)	reduces methy glyoxal and 2,5-diketogluconate	00248	NP_417485
YajO	putative NAD(P)H-dependent xylose reductase	00248	NP_414953
YghZ	reduces methy glyoxal	00248	NP_417474
Tas	Multicopy expression suppresses prephenate dehydrogenase defect	00248	NP_417311
YdhF	predicted oxidoreductase	00248	YP_025305
Ydbc	predicted oxidoreductase	00248	NP_415924
Short-Chain Dehydrogenases			
ybbO	short chain dehydrogenase	pfam00106	NP_415026
yohF	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	pfam00106	NP_416641
YciK	short chain dehydrogenase	pfam00106	NP_415787
YgfF	predicted NAD(P)-binding oxidoreductase with NAD(P)-binding Rossmann-fold domain	pfam00106	NP_417378
YghA	oxidoreductase	pfam00106	NP_417476
YgiI	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain	pfam00106	NP_418670
YdfG	putative oxidoreductase, L-allo-threonine dehydrogenase, NAD(P)-binding	pfam00106	NP_416057
YgcW	putative oxidoreductase, predicted deoxygluconate dehydrogenase	pfam00106	NP_417254
UcpA	predicted acetoin dehydrogenase (diacetyl reductase)	pfam00106	NP_416921
EntA	2,3-Dihydro-2,3-dihydroxybenzoate dehydrogenase	pfam00106	NP_415128
FolM	dihydrofolate reductase, THF biosynthesis	pfam00106	NP_416123
HdhA	7-alpha-hydroxysteroid dehydrogenase	pfam00106	NP_416136
HcaB	2,3-dihydroxy-2,3-dihydrophenylproponate dehydrogenase, (aromatization)	pfam00106	NP_417036
SrlD	sorbitol-6-phosphate 2-dehydrogenase, glucitol-6-phosphate dehydrogenase	pfam00106	NP_417185
KduD	2-deoxy-D-glucuronate 3-dehydrogenase	pfam00106	NP_417319
IdnO	gluconate 5-dehydrogenase, 5-keto-D-glucuronate 5-reductase	pfam00106	NP_418687
FabG	3-oxoacyl-[acyl-carrier-protein] reductase	pfam00106	NP_415611
FabI	enoyl-(acyl carrier protein) reductase	pfam00106	NP_415804
YdJA	predicted oxidoreductase (FAD dep.)	pfam00881	NP_416279

FIG. 10B

ygjB (NP 418690)

Nucleotide sequence (SEQ ID NO:93)

1 atgtcgatga taaaagcta tgcccaaaaa gaagcgggcg gcaaactgga agtttatgag  
61 tacgatcccg gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcggatc  
121 tgcattccg atctgtcgat gatcgataac gaatggggat tttcacaata tccgctggtt  
181 gccccggcatg aggtgattgg gcgcgtggg gcaactcgga gcccgcga ggataaagggt  
241 ttgcaggtcg gtcagcgtgt cgggattggc tggacggcgc gtagctgtgg tcactgcac  
301 gcctgttatta gcggtaatca gatcaactgc gagcaagggt cggtccgac gattatgaat  
361 cgcggtgtgc ttgcggagaa gttgcgtgcg gactggcaat gggtgattcc actgccagaa  
421 aatattgata tcgagtcgc cggggcgtg ttgtgcggcg gtatcacggt cttaaacc  
481 ctgttgcgtc accatatac ac tgctaccagc cgcgttgggg taattggat tggcggtctg  
541 gggcatatcg ctataaaact tctgcacgc atgggatgc aggtgacacgc cttagttct  
601 aatccggcga aagagcagga agtgcgtgcg atgggtgcgg ataaagtggt gaatagccgc  
661 gatccgcagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac  
721 gtcagcctcg actggcagcc ctatttgag ggcgtgaccc atggcggtaa ttccatagc  
781 gtcgggtgcgg ttctcacgcc gctgtctgtt ccggcccttta cgttaattgc gggcgatcgc  
841 agcgtctctg gttctgtac cggcacgcct tatgagctgc gtaagctgat gcgttttgc  
901 gccccgcaga aggttgcggc gaccaccgaa ctgttcccgat tgtcggaaat taacgacgc  
961 atccaqcatq tgccqcaqcg taaggcqctt accgcqcttq tgttgaaaat cgatccat

Amino acid sequence (SEQ ID NO:94)

```

1 msmiksyaak eageleveye ydpge1rpqd vevqvdyfcgi chsdlsmidn ewgfsqyplv
61 aghevigrvv algsaaqdkg lqvgqrvgig wtarscghcd acisgnqinc eggavptimm
121 rggfaeklra dwqwvpiplre nidiessagpl lcggitvfkpl 1lmhhitats rvvgigiggl
181 ghiaikllha mgcevtafss npakeqevla mgadkvvnsl dpqalkalag qfdliintvn
241 vsldwqpyfe altyggnfht vgavlplsv paftliagdr svsgsatgtp yelrklmrfa
301 arskvaptte lfpmksinda ighvrdrdgkar yrvvvlkadf

```

yahK (NP 414859)

Nucleotide sequence (SEQ\_ID\_NO:95)

```
1 atgaagatca aagctgttgg tgcattttcc gctaaacaac cacttgaacc gatggatata  
61 acccggcgta aaccgggacc gaatgtatgc aaaatcgaaa tcgccttactg tggcggttgc  
121 cattccgatc tccaccaggc ccgttccgag tggcggttggc cggtttaccc ctgcgtgcc  
181 ggtcatgaaa ttgtggggcg tttgttagcc gttgtgtatc aggttagaaaa atatgcgc  
241 ggcgatctgg tcgggtgtcg ctgcattgtc gacagtgtatc aacattgcga agagtgtgaa  
301 gacgggttgg aaaactactg tgatcacatg accggcacct ataactcgcc gacccggac  
361 gaaccgggccc atactctgg cggtactcta caacagatcg tcgttcatga gcgatatgtt  
421 ctgcgtattc gtccacccca agagcagctg gcccgggtgg ctcctttgtt gtgtgcagg  
481 atcaccacgt attcgccgct acgtcactgg caggccggc cgggtaaaaa agtgggcgtg  
541 gtcggcatcg gcggctctgg acatatgggg attaagctgg cccacgcgt gggggcacat  
601 gtgggtggcat ttaccacttc tgaggcaaaa cgcgaagcgg caaaagccct gggggcccat  
661 gaagtgtta actcacgca tgccgatgag atggcggctc atctgaagag tttcgatttc  
721 atttgtata cagtagctgc gccacataat ctcgcacgtt ttaccaccc tgcgttgtt  
781 gatggcacca tgacgttgtt tggtgcgcct ggcacaccgc ataaatcgcc ggaagtttc  
841 aacctgtatca tgaaacggc tgccatagcc gtttctatga ttggcggtat tccagaaact  
901 caggagatgc tcgatcccc ccggcaacat ggcacatgtgg ctgtatata gatgttccgg  
961 gccgatcaaa ttaatgaagc ctatggcga atgctgcgcg gtgtatgtaa atatcgaaaa  
1021 ttatcgata atcgacact aacagactgtat
```

Amino acid sequence (SEQ\_ID\_NO:96)

1 mkikavgays akqplemdi trrepgpndv kieiaycgvc hsdlhqvrse wagtvypcvp  
61 gheivgrvva vgdqvekyap gdlvgvgciv dsckhceece dglenycdhm tgtnsynptpd  
121 epghtllggys qqivvheryv lrirhpqeql aavapllcag ittysplrhw qagpgkkvgv  
181 vgigglghmg iklahamgah vvafttseak reaakalgad evvnsrnade maahlksfdf  
241 ilntyaaiphn lddfttlkr dgtmtlygap atphkspevf nlimkrraia qsmiqgipet

**FIG. 10C**

301 qemldfcaeh givadiemir adqineayer mlrgdvkyrf vidnrtltd

**adhP (NP\_415995)****Nucleotide sequence (SEQ ID NO:97)**

```

1 atgaaggctg cagttgttac gaaggatcat catgttgacg ttacgtataa aacactgcgc
61 tcactgaaac atggcgaagc cctgctgaaa atggagtgtt gtgggttatg tcataccgat
121 cttcatgtta agaatggcg tttgggtac aaaaccggcg taattctggg ccatgaaggc
181 atcgggtgg tggcagaagt gggtccagggt gtcacacctat taaaaccagg cgatcgtgcc
241 agcgtggcg ggttctacga aggatgcggt cattgcgaat actgttaacag tgtaacgaa
301 acgcctctgcc gttcaggtaa aaatgcccga tacagcgttg atggcgggat ggcggaaagag
361 tgcatcggtt tcgcccatta cgcggtaaaa gtgccagatg gtctggactc ggcggcggcc
421 agcagcatta cctgtgcggg agtaccacc tacaaagccg ttaagctgtc aaaaattcgt
481 ccagggcagt ggattgttat ctacggctt ggcggcttgg gtaacctcgc cctycaatac
541 gcgaagaatg tctttaacgc caaagtgatc gccattgtat tcaatgtatga gcagttaaaa
601 ctggcaaccg aaatggcg agatttagcg attaactcac acaccgaaga cgccgcggaaa
661 atttgtcagg agaaaaactgg tggcgtcac gctgcgggtt taacagcggt agctaaagct
721 gcgtttaact cggcaggta tgctgtccgt gcaggcggc gtgttggc tgcggctcta
781 cccggcggagt ctatgagcct gatatcccc cgtttgtgc tggatggat tgaagtggc
841 ggttcgctgg tcggcacgcg ccaggattta actgaagcct tccagttgc cccgaagggt
901 aaagtgggtgc cgaaaatcgc cctgcgtccg ttacggaca tcaacaccat cttactgag
961 atgaaagaag gcaaaatccg tggccgcata gttttttttt tccgtcacta a

```

**Amino acid sequence (SEQ ID NO:98)**

```

1 mkaavvtdh hvdvtyktlr slkhgeallk meccgvchtd lhvkngdfgd ktgvilghes
61 igvvaevgpg vtslkpgdra svawfyegcg hceycnsne tlcrsvknag ysvdggmaee
121 civvadyavk vpdglldsaaa ssitcavtt ykavklskir pggwiaiygl gglgnlalqy
181 aknvfnakvi aidvnndeqlk latemadla inshtedaak ivqektgah aavvtavaka
241 afnsavdavr aggrvvavgl ppesmsldip rlvdgievv gslvgtrqdl teafqfaaeg
301 kvvpkvalrp ladintifte meegkirgrm vidfrh

```

**ydjL (NP\_416290)****Nucleotide sequence (SEQ ID NO:99)**

```

1 atgaaagcac tggctcggtt tggcaaggcc tttggcggct acaagatgtat tgatgtccca
61 caacccatgt gtggcccgga agatgtatgtt attgaaatattt aagccgcggc aatctgcggc
121 gcagacatga agcactacaa tgtcgatagc ggttctgtat agtttaactc tatccgcggc
181 catgagttcg cagggttat tgccgcagggtt ggtaaaaag tcaaagactg gaaagtgggg
241 caacgcgtcg tattcgatataa cagcggtcac gtttgcgggtt tttgtccggc ctgtgaacaa
301 ggtgattttc tggatgttac agaaaaagta aacctggc tggataataa tacctggggc
361 ggtggttttt ccaaataattt tctggttccct ggtgaaattt tcaaaattca tcgtcatgc
421 ttgtggaaa tccctgtatgg tggatgtt gaggacgcag ccgtacttga ccctatctgt
481 aatgcctaca aatccatgc gcagcaatcg aatttccttc ctggtcagga tggatgtcgtc
541 atccggcactg gcccactcgg gctgttctcc gtacaaatgg cgcaattat gggggcggta
601 aatatcgtcg tggatgtt gcaagaagat gtggcggtcc gttcccggt tgcaaaagaa
661 ctgggtgcga cggcagtagt aaatggttt accgaagatg tggtggcgcg ctggcagca
721 atttgtggca aagacaatct gggactgggtt attgaatgtt ccggtgccaa tatcgactg
781 aaacaagcca tggaaatgtt ccggccggaaac gggaaatgg tacgcgttgg aatgggcttc
841 aaacccctttt atttctcgat taatgacattt accgcctgg aaaaaagcat cattggggcat
901 atggccatctt actccaccc atggcgttaac gttatcgaggc tattagccag cggcgtatc
961 aaagtcaaac cgtatgtatc gcatcgatc ggcctgtcgc aatggcgcga agggtttgat
1021 gcatggtcg ataaaaccgc aatcaaagtg atcatgactt acgactttga tgaataa

```

**Amino acid sequence (SEQ ID NO:100)**

```

1 mkalarfkgka fggkykmidvp qpmpcgpeddvv ieikaaaicg admkhynvds gsdefnsirg
61 hefacgiaqv gekvkdwkvg qrvvsdnsgh vcvpcpaceq gdflcctekv nlglndntwg
121 ggfkskyclvp geilkahrha lweipdgvdv edaavldpic nayksiaqqks kflpgqdvvv
181 igtgplglfs vqmarimgav nivvvglqed vavrvfpake lgatavvngs tedvvarcqq
241 icgkdnlgkv iecsganial kqaiemrpn gevrvvgmgf kpldfsindi tawnksiiigh

```

**FIG. 10D**

301 maydstswrn airllasgai kvkpmithri glsqwregfd amvdktaikv imtydfde

**ydjJ (NP 416288)****Nucleotide sequence (SEQ ID NO:101)**

```

1 atgaaaaatt caaaagcaat attgcaggtg ccgggcacaa tgaaaattat ttcagcagaa
61 ataccagtgc ctaaagaaga tgaagtttg attaaagttag aatatgtcg tatttgcgtt
121 tcagatgtac atggtttga atcaggcccg tttattccgc ctaaagaccc aaatcaagaa
181 attggcctgg gtcataatgc atcggggacg gttgtggctg tgggaagccg cgtgcgcaaa
241 tttaaaccgg gggatcggtt aaatatcgaa cctggcggtc cttgcggta ctgtcggtac
301 tgtctggaa gcaaataataa catctgcggc gacgttgatt ttatggcgac acaacccaaac
361 taccgcggcg cattaacgc cttatctgtgt catccggaga gctttactta caaactgccc
421 gacaatatgg acacgatgga agggcgctg gtggagcctg ccgcagtcgg gatgcgtgcc
481 gcgatgctgg cagatgttta accgggttaag aagataatta ttctggagc aggttgttatt
541 gtttgcgttgc cgttgcggc gtgcggatgc ctggggacaa cggaaattgc cgtcggttgc
601 gtgcgtggaa aacgtctggc aatggcgaa cagcttgcgtt cgacagtgg tattaacggc
661 gcaaaaagaag acactattgc acgcgtgtcgtt caatttaccg aagacatggg cgccggatatt
721 gttttcgaaa cagcggttgc tgcgttgcacc gttaaacagg caccttatct ggtaatgcgc
781 ggcggtaaaa ttatgtttttt tggactgtt cccggcgatt cggcaatcaa ttccctcaaaa
841 atcaatgcgc aagtcaatcc gagacggta ttccgctat ccaatcggtt tccgggtcacc
901 attgcgttgc ttatccgttgc ggcatttcgtt gtgcgttgc tgggtacgca tattttacgt
961 tatcggttgc tacaacaggc atttgcgttgc tcagtttaca acaaaacgcga cattattaa
1021 ggcgttattaa aaatttgcga tttaa

```

**Amino acid sequence (SEQ ID NO:102)**

```

1 mknksailqv pgmtkiisa ipvpkedevl ikveyvgicg sdvhgfesgp fippkdpnqe
61 iglghecatg vvavgsrvrk fkpgdrvnie pgvpcghcry clegkynicp dvdfmatqpn
121 yrgalhylc hpesftyklp dnmdtmegal vepaavgmha amladvkpgk kiiilgagci
181 glmtlqackc lgateiavvd vlekrlamac qlgatvvting akedtiarcq qftedmgadi
241 vfetagsavt vkqapylvrm ggkimivgtv pgdsainflk inrevtiqtv fryanrypvt
301 ieaiissgrfd vksmvthiyd yrdvqqafee svnnkrdiik gvikisd

```

**idnD (NP 418688)****Nucleotide sequence (SEQ ID NO:103)**

```

1 atgcaagtga aaacacatgc ctgcgttgc ttgcggcaaga aaactgttgc cgttaccgag
61 cagacgatag attggaaataa taatggaaaca ttgttacaaa taacccgagg tggaaatttgc
121 gtttccgtt tacatttata tcaggaagga aaagttagtta atttcatgtt aaaggcaccg
181 atgggtttag gtcataatgc tttccgtttaa gtttccgttgc gtcactgc aaatgttgc
241 gaaggccaaat ccgttccgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
301 gaacataacg agaatcgttgc tacagatgttgc ttgcgttgc ttgcgttgc ttgcgttgc
361 catgttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
421 ccggccaaat ctgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
481 gcccacatc aggcggcgaa tttccgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
541 attggcttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
601 gatgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
661 ccacaaaacg acgacatggc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
721 gaagtgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
781 gtaatgggttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
841 ggtaaggaga ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
901 tcatggctgg cgaatggcgatc tatcaatccca ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
961 actgacatggc acgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc
1021 ttgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc ttgcgttgc

```

**Amino acid sequence (SEQ ID NO:104)**

```

1 mgvktqscvv agkkktvavt qtidwnnnngt lvqitrggic gsdlhyyyqeg kvgnfmikap
61 mvlghevigg vihsdssselh egqtvainps kpcghckyci ehnencqcdm rffgsamyfp
121 hvdggftryk mvetsqcvcv pakadekvma faeplavaih aahqagelqq krvfisgvvp
181 igclivsavk tlgaeeivca dvsprslslg kemgadvlvn pqnddmldhwk aekgyfdvsf
241 evsghpssvn tclevtrarg vmvqvqmgga maefpmmtli gkeislrgsf rftsefnlav

```

FIG. 10E

301 swlangvinp lpllsaeypf tdlealrfa qdktqaakvq lvf

tdh (NP 418073)

Nucleotide sequence (SEQ ID NO:105)

1 atgaaaagcgt tatccaaact gaaagcgaa gagggcatct ggatgaccga ctttcgttgc  
61 ccggaaactcg ggcataacga tctgtgttatt aaaatccgtt aaacagccat ctgcgggact  
121 gacgttcaca tctataactg ggatgagtgg tcgcaaaaaa ccatccggt gccgtatggc  
181 gtgggccatg aatatgtcgg tgaagtggta ggtattggtc aggaagtgaa aggcttcaag  
241 atcggcgatc gcgtttctgg cgaaggccat atcacctgtt gtcattgccc caactgtcgt  
301 ggtggtcgtt cccatttgcg ccgcacacg ataggcggtt gtgttaatcg cccgggctgc  
361 ttgcgcgaat atctggtgat cccggcattt aacgccttca aaatccccga caatatttcc  
421 gatgacttag ccgcatttt tgatcccttc ggttaacgcgc tgcataccgc gctgtcgttt  
481 gatctggtgg gcgaagatgt gctgtttct ggtgcaggcc cgattggtat tatggcagcg  
541 gcgggtggcga aacacgttgg tgcaacgcattt gtgggtatca ctgatgttaa cgaataccgc  
601 ctgtagctgg cgcgtaaaaat gggtatcacc cgtgcgggtt acgtcgccaa agaaaatctc  
661 aatgacgtga tggcggagtt aggcatgacc gaagggtttt atgtcggtt ggaatgtcc  
721 ggtgcggccgc cagcgtttcg taccatgtt gacacatgtt atcacggcgg ccgtattgcg  
781 atgctggta ttccggcgtc tgatatgtct atcgactggc ccaaagtgtat cttaaaggc  
841 ttgttcattt aaggatattt cggcgtgag atgtttgaaa cctggtacaa gatggcggcg  
901 ctgattcgt ctggcctcga tctttcggcgtt atcattaccgc atcgtttctc tatcgatgtat  
961 ttccagaagg gctttgacgc tatcggttgc ggccagtcgg gaaaagttat tctgagctgg  
1021 gattaa

Amino acid sequence (SEQ\_ID\_NO:106)

1 mkalsklkae egiwmtdvpv pelghndllicirktaicgt dvhiyndew sqktipvmw  
61 vgheyvgcgv gigqevkgfk igdrvsgegh itcghcrnrc ggrthlcrrnt igvgnrpgc  
121 faeylvipaf nafkipdnis ddlaaifdpf gnavhtalsf dlvgedvlvs gagpigimaa  
181 avakhvgarn vvitdvneyr lelarkmgit ravnvakenl ndvmaelgmt egfdvglems  
241 gappafrml dtmmhggria mlgippdms idwtkvifkg lfikgiygre mfetwykmaa  
301 liqsqldsp iithrfsidd fqkgfdamrs qgsqkvilsw d

yjjN (NP 418778)

Nucleotide sequence (SEQ ID NO:107)

1 atgtctacga tgaatgttta aatttgccag cagccgaaag aattagtctg gaaacaacgc  
61 gagataccta ttccgggtga caatgaagca ttaataaaaaa ttaagtctgt cgggatttgc  
121 ggtaccgata ttcatgcctg gggtgaaaat caaccattt ttagttatcc acgtgtttta  
181 ggcctatgaaa tatgtgggaa gattgttggg ctgggtaaaa atattgctga tcttaaaaat  
241 ggtcagcaag ttgctgtat cccttatgtt gcctgtcago aatgcccggc gtgtaaaaggc  
301 gggcgatcca attgctgtga aaaaatttca gtcatggcg tgcatacgga tgccgggttt  
361 agtgagttt tgagcgtgcc ggtggcgaac attttgcggc cagacggat tgaccccgac  
421 gccgcagcat tgattgaacc ttgcgtatt agcgtcatg cggtgcgtcg cgccggcatt  
481 gctccggcg agcagggtct ggtgtcggg gcggggccaa tcggcttggg cgccggcgca  
541 atcgctaaag ccgatggcg acagggtgt gtggcgatca ccagtccggc ggcggctgaa  
601 catgtggcaa cgcgtcttga attacktttta ctggaccggcgt cagccgaaaga ttggacccgc  
661 cagctactggg cgcagtttgg tggtcgtcgc ggcgcgaaaag tgatgcacgc gacaggtaat  
721 caacatgcga tgaataaacac cgtgtatgg attcgtcaacg gccgcacggg ggtatttgc  
781 ggctgttta aagggtgatgt gcagttctcc gatccggaaat tccataaaaaa agaaacgcacg  
841 atgatggcga gcccacgcg cacccggaa gattttgcata aagtccgtcg actgatggcg  
901 gaaggaaaaaa tcaactgtca catgttta accccatcgat tccgtcgactgatggcg  
961 gaaacctacg agcgcgtatgt gattaacaat cgtgatgttta taaaaggcgt aattacttcc  
1021 tga

Amino acid sequence (SEQ ID NO:108)

1 mstmnvlicq qpkelvwkqr eipipgdnea likiksvgic gtdihawgn qpffsyprvl  
61 gheicievlg kgniadlkn gqqvavipvy accqqcpacks grtnceekis vigvhqdggf  
121 seylsvpvan ilpnadgidpq aaaliepfai sahavrraai apgeqvllvv agpiglgaaa  
181 iakadqaqvvy vadtsparre hyatrlepl ldpesaedfa qlraqfqgsl aqkvidatqn

**FIG. 10F**

241 qhamnnntvnl irhggtvvfv glfkgelqfs dpefhkkett mmgsrnatpe dfakovgrlma  
 301 egkitadmml thrypfatla etyerdvinn relikgvitf

**rspB (NP 416097)****Nucleotide sequence (SEQ ID NO:109)**

```

1 atgaaaagca tattaaattga aaaaccgaat caactggcga ttgtcgaacg taaaatacc
 61 acccccgttag cgggtgaagt acgagtaaaa gtgaaacttg ccggaaatttg tggttcagat
121 agccatattt atcgtggca taatctttt gcgaaatatc cgcgcgtcat tggtcatgaa
181 ttctttggcg tcattgtgc agtgggtgaa ggcgtggaa gcgcaggagt cggtaacgt
241 gttgctgtcg atccgttgtt cagctgtggg cattgtatc cgtgcttat aggtaaaccg
301 aacgtttgcg cgcacactggc tgttatttagt gtgcacgctg acggtggtt cagtgatata
361 gccgtggttc cggcaaaaaaa tgcgtggaaa attcctgaag cagtggccga tcaatatgcg
421 gtaatgatcg aaccttttac cattgcggct aacgtaaccc gacatggtca accgactgaa
481 aatgataccg ttctggttta tggtgcgggt ccaatcggcc tgacgatcgt tcaggttata
541 aaaggcgtct ataacgttta aaalglgatt gtlgcccgtc gcattgtatc acgactggaa
601 aaagcqaaag agagcggggc tgactggcg attaataaca gccagacacc gcttggcgag
661 attttactg aaaaaggcat caagccgaca ttaattatcg atgcggcttgc tcattttct
721 atccctgaaag aggccgtaaac gctggcttcc cagccggcac gtattgtatt gatgggggtt
781 tccagtgaaac cgtctgttgc gattcagcaa ggaattaccg gaaaagaact ctctattttc
841 tcttcacgt taaatgaaaaaa taaatcccg atcgttatacg actggtaag taaagggtta
901 attaaaccag aaaaattat taccatatacg ttgtattcc agcatgttgc tgatgccatt
961 agtttatttg aacaggatca aaaacattgc tgcaaagtct tactcactt ttctgataaa

```

**Amino acid sequence (SEQ ID NO:110)**

```

1 mksiliekpn qlaivereip tpsagevrk vklagicgsd shiyrgnlpf akypervighe
 61 ffgvidavge gvesarvger vavdpvvscg hcypcsigkp nvcttlavlg vhadggfsey
121 avvpaknawk ipeavadqya vmieptfiaa nvtghqgpte ndtvlyvgag pigltivql
181 kgvynvnkvni vadriverle kakesgadwa innsqtplge iftekikpt liidaachps
241 ilkeavtlas paarivlmgf ssepseviqq gitgkelsif ssrlnankfp ividwlskgl
301 ikpeklitht fdfqhvhadai slfeqdqkhc ckvlltfse

```

**gatD (NP 416594)****Nucleotide sequence (SEQ ID NO:111)**

```

1 atgaaatcag tggtaatgt tactgtatggt atcgtgcgcg ttgcagaaag cgtcatttc
 61 gaaattaaac atcaggatga ggtgcgggtt aaaaattgcctt gctcgggtt atgtggttcc
121 gatttaccca ggatatttaa aaatggtca cattattatc caataacgtt agggcatgaa
181 tttagcggctt atattgtgc ggtggatcc ggtgttgatg atttacaccc tggcgatgc
241 gttgcctgtt tgcgttattt accctgtttt acttgtccag agtgtctgaa aggttttat
301 tcccagtgcg caaaatatga ttttattggc tgcggcggtt atggtgatt tgctgatata
361 attgtcgtaa agcgaaaaaa tgcgtttgtt ctacccacgg atatgcctt tgaggatggg
421 gcttttattt agccgattac cgttggcttgc catgcttttcc atttagcgca aggtgtgag
481 aataaaaaacg ttattattat tggtgccgaa accattggcc tgctggccat tcagtcgct
541 gtcgcgctgg gagcaaaagag tgcgtacggcg atcgcacatta gttcagaaaaa actggcactg
601 gcaaaatctt tcggtgcgtt gcaaacattt aacagttagcg aatagagcgc gcccacaaatg
661 caagcggtt tacgcgaact ggcgtttatc cagtttatcc tcgagacggc tggcgatcc
721 caaactgtcg aactggcggtt agagattgcc ggtcctcatg cccaaactggc gctgggtggc
781 acgttgcato aggtatctgc tttaacatcg gcaacgtttt gcaaaatattt gctaaagag
841 ctgacgggtt tcggcagggtt gatgaaacttgc tccagccctt ggccggggca ggagtgggaa
901 acggcgagcc ggttgcgttgc agaacgttgc ttaagccttgc agccattaat cgctcaccgt
961 ggaagcttgc aaagcttgcg ccaggcggtt cgtgacatcg ctgcgtatgc tatgcggggc
1021 aaagtgttgc tcattccctg a

```

**Amino acid sequence (SEQ ID NO:112)**

```

1 mksvvndtdg ivrvaesvip eikhqdevrv kiassglcgs dlprifknqa hyypitlghe
 61 fsgyidavgs gvddlhpgda vacvplpcf tcpeclkgfy sqcakydfig srrdggfaey
121 ivvkrknvfa lptdmpiedg afiepitvgl hafhlaqgce nknviiigag tigllaiqca
181 valgaksvta idisseklal aksfgamqtf nssemsapqm qsvlrelnfn qilettagvp

```

**FIG. 10G**

241 qtvelaveia gphaqlalvg tlhqdhlts atfgkilrke ltvigswmny sspwpgqewe  
 301 tasrllterk lslepliah rgsfesfaqav rdiarnampg kvllip

**yphC (NP 417040)****Nucleotide sequence (SEQ ID NO:113)**

```

1 atgaaaacga tgctggcagc ttatttacca ggaaattcga ccgtcgatct gcgggaagtt
 61 gcgggtccga cgcggggat taaccaggta ctgatcaaaa taaaatcctc cgggatttgc
121 ggaaggcatg tccactatac ctatcatcaa caccgtcca cagcggcggc acccgataaaa
181 cccgttatacc agggcttat caacggtcat gaaccgtgc ggcagattgt ggcgtatgggg
241 caaggctgcc gccattttaa agagggcgc acggcgtctgg tgtatcacat ttctggctgt
301 gtttttgcg cgaactgccc tcgggtttt cctatttctt gtactggcga aggaaaagcg
361 gcttacggct ggcagegtga cggcggatcat gccgaactact tgcggcggaa agaaaaagat
421 ctgatcctc tgccggatgc gtcgatctac gaagatgggt cgtttatcat ttgcggcgtt
481 ggtacagcgt atgaaggaaat tttgcgcgc gaagtttccg cgtgtataaa cgtgctgggt
541 gtcggctctgg ggccagtcgg catgatggcg atgatgtctgg cggcggcggaa
601 cggatcatcg gcgttgatcat gtcggcggaa cgtctggcga tggcaaaaca gtttaggggtg
661 atggatcacy gctatttagc aaccaccgaa ggtctggcgc agattatcgc cgaactcacc
721 cacggtggcg cggatgttgc gtcgatgttgc tccggtaatg cccgaggcgtc cttgctggca
781 ctgcaatcca cccgtactg gggacgggtg gtttacattt gtgaaaccgg aaaaatggaa
841 ttcgaggtca gcccgcgatct gatgcaccat caacggcggaa ttatcgctc ctgggtgacc
901 agtctgttcc atatggaaaat atgcgcctt gatctgacgg actggaaagct gtggccgcgt
961 aacgcattt cccatcgctt ctcgctggaa caggcagggtg atgcctatgc gtcgtatggcg
1021 agcggcaaat gcgggaaagt tgtgattaac ttccggatt aa

```

**Amino acid sequence (SEQ ID NO:114)**

```

1 mktmlaaylp gnstvdllrev avptpginqv likmkssgic gsdvhyiyhq hrataapdk
 61 plyqqfinh epcqqivamg qgcrhfkegd rvlvyhisgc gfcpncrrgf pisctegk
121 aygwqrddgh aeyllaeeekd lillpdalsy edgafiscgv gtayegilrg evsgsdnvlv
181 vglgpvgmma mmlakgrgak riigvdmlpe rlamakqlgv mdhgylatte glpqiaelt
241 hggadvaldc sgnagagrll aqstadvgrv vyigetgkve fevsadlmhh qrriigservt
301 slfhmekah dltdwklwpr naithrfsle qagdayalma sgkcgkvvin fpd

```

**yhdH (NP 417719)****Nucleotide sequence (SEQ ID NO:115)**

```

1 atgcaggcgt tactttaga acagcaggac ggcaaaactc tcgcattcgtt acagactctg
 61 gacgaaagtc gcctgcccggaa gggcgatgtc acggcgtatgtt tcactggc gggctgaa
121 tataaaatgcg cgcgtggat taccggtaag ggaaaaatca tccgtatattt tccgtatgatt
181 cctgggatcg attttggcg aactgtacgc accagcgaag atccgcgtt tcattccgg
241 caggagggtgt tactcactgg ctggggcggtt ggtggaaacc actgggggtgg gtcggcggag
301 caggcgcgag tggaaagggtga ctggctgggtt gccatggcgc aaggggctggaa cgcgcgtaaa
361 gcaatgatta tcggtactgc cgggtttacc gccatgtgtt gtgtatggc gtcggaaat
421 gcccggatcc gcccgcagga cggggagatt gtcgtacgg gtcggcgtt tggcgtccgc
481 agtacccgcg tggcgctgtc gcataatgtt ggttattcagg tcgttgcgtt ttccggcgtc
541 gaaagtaccc atgaatatct gaaaatgtt ggtgtatgtt gtttctccc tcgtatgtt
601 tttggcgtt cccgtcctt gggaaacaa gtcgtggctg gggcaatttga caccgttggc
661 gacaaagtgc tggcaaaatgtt gtcgtggctg gggcaatttga caccgttggc
661 gacaaagtgc tggcaaaatgtt gtcgtggctg gggcaatttga caccgttggc
721 ggtctggcggtt gttttttac tctgcaccc acggcgtatgc cattttatctt gtcgtatgtc
781 cgttgcggatgggtggatttactt gtcgtggctg gggcaatttga caccgttggc
841 cgttgcggatgggtggatttactt gtcgtggctg gggcaatttga caccgttggc
901 tcagaggcactt cgttgcggatgggtggatttactt gtcgtggctg gggcaatttga caccgttggc
961 gtgaagggtt actaa

```

**Amino acid sequence (SEQ ID NO:116)**

```

1 mqallleqqd gktlasvqt1 desrlpegdv tvdvhwssl1 ykdala1tgk gkiirnfpmi
 61 pgidfragtv1 tsedprfhaq qevl1tgwgv genhwglae qarvkgdwl1v ampqgldark
121 ami1gtagft amlcvmaled aqvrpqd1gei vvtgasggvg stavallh1l gyqvavasgr
181 estheylks1 gasrvlprdc facsrplekq vwagaidtv1 dkvlakvlaq mnnyggcvaac

```

**FIG. 10H**

241 glaggftlpt tvmpfilrvn rlqgvdsvmt pperraqawq rlvadlpesf ytqaakeisl  
 301 seapnfiaeai innqiqgrtl vkvn

**ycjQ (NP 415829)****Nucleotide sequence (SEQ ID NO:117)**

1 atgaaaaagt tagtagccac agcaccgcgt gttgctgcgc tggttgagta tgaagatcg  
 61 gcgatttttag ctaatgaagt gaagatccgc gtgcgttgc gcgcaccgaa acacggAACG  
 121 gaagtggtcg acttccgcgc cgccagcccc ttattgtatg aagactttaa cggcgaatgg  
 181 cagatgttca ctccgcgtcc ggcagatgcg cgcgcggca ttgagtttgg caaattccag  
 241 ctggcaaca tggtggttgg cgacattatc gagtgcggca ggcacgttac cgactacgc  
 301 gtggcgcaca gcgtatgcgg ctacggccc ctctccgaga cggtcatcat taacgcgtg  
 361 aataactaca agctgcgc当地 aatggcccaa ggcaacttctt ggaaaaaacgc cgtctgtac  
 421 gacccggcgc当地 agtttgc当地 gagcggcgtg cgcgc当地 acgtgcgc当地 aggggatttt  
 481 gtggtggttgg tagggcttgg cgcaatcggt caaattgcca tccaaacttggc taaacgc当地  
 541 ggcgc当地tctg tggtgattgg cgtcgatctt atcgc当地tccatc gctgtatat tgccgc当地  
 601 cacggcgc当地 atttctgc当地 taacccatc ggcaactgatg taggtaaaga gatcaaaaacg  
 661 ctgacccggca agcagggtgc cgtatgtatt atcgaaacca ggc当地tacgc cga当地gc当地  
 721 caatcgccgc当地 tccgc当地tctt ggcttatggc ggc当地tacatc cctatgtgc当地 gtttgc当地  
 781 ccgttgc当地 aaggttt当地 aacttggc当地 gaaagc当地tacatc tcaataacgc caaaatttgc当地  
 841 ttctctgc当地 cgtgc当地cgc当地 accgaacccg gattatccgc gctggagccg caagc当地tac  
 901 gaagaaacct gttgggaact gctgatgaaac ggttatctca attgcaaga tttatcgac  
 961 ccggtggtga ccttgc当地 tagccggaa agctatatgc agtatgtcga ccagcatccg  
 1021 gaacagagca tcaaaaatggg cgttacgtt当地 taa

**Amino acid sequence (SEQ ID NO:118)**

1 mkkvlvatapr vaalveyedr ailanevkir vrfgapkhgt evvdfraasp fidedfngew  
 61 qmftprpada prgiefgkfq lgnmvvgdii ecgsdvtaya vgdsvcgygp lsetviinav  
 121 nnyklrkmpq gsswknnavcy dpaqfamsgv rdanrvvgdf vvvvglgaig qiaiqlakra  
 181 gasvvigvdp iahrcdiarr hgadfclnpi gtdvgkeikt ltgkqgadvi ietsgyadal  
 241 qsalrglayg gtisyyafak pfaefnlgr eahfnnakiv fsracsepnp dyprwsrkri  
 301 eetcwellmn gylncedlid pvvtfanspe symqyvdqhp eqsikmgvtf

**yncB (NP 415966)****Nucleotide sequence (SEQ ID NO:119)**

1 atggggcaac aaaaggcagcg taatcgacgt tgggttctgg cctcgctcc acatggcgca  
 61 cctgttccgg agaatttccg tcttgaagaa gatgatgtcg ccacacccgg tgaaggacag  
 121 gtgttactgc gcacagtttta tttgtccctg gaccegtata tgcgtggacg tatgagcgat  
 181 gagccatctt attcaccgc当地 tggatatt ggc当地ggc当地 tggc当地ggc当地 tacgtgagc  
 241 cgtgtcg当地 agtcaatca tcctgattat cagtctggcg actgggtgt gggctacag  
 301 ggatggcaag actatgacat atccagttgt gatgatctgg tggaaacttgg cgatcatccg  
 361 caaaaatccat cgtggc当地 ggggtgtcta gggatgccag gcttaccgc ttatatggc  
 421 ctactggata tcggc当地tgc当地 taaagaggcc gaaacgttgg tggtagtgc ggc当地acagga  
 481 ccagtggggcc cgacggc当地 gcaaatcgcc aaaaacttgg gttc当地gaggt ggtgggggta  
 541 gccggc当地 cggaaaaatccat cgc当地atgtt accggagggtg taggttccg tgggttgc当地  
 601 gatcaccacg cggatgattt tggcaacaat ctggcgaaag cgtgccc当地 aggtattgtat  
 661 atctattatg aaaaatgggg cggtaaggta ttgc当地atgc当地 tgctaccgtt acttaataaca  
 721 tctgc当地gc当地 tcccgcttgc cggatgttgc acgatgttgc gctaccacc  
 781 gttccggatc gtttacctt gttgatggct acagtgc当地 aaaaacgtat tggcttgc当地  
 841 gtttttatta tcgctc当地 gggatgttgc acgatgttgc agtttccatc gggatgggg  
 901 caatgggtga aagaggatata aatccactac cgc当地agaaa ttactgc当地 gtttagagaat  
 961 gccc当地acaga cgttacccat cctgtc当地 ggtaaaaact tggcaaaatg ggtgatccgc  
 1021 gttgggggtt当地 atgattaa

**Amino acid sequence (SEQ ID NO:120)**

1 mgqqkqrnrr wvlasrpgha pvpnenfrlee ddvatpgegg vllrtvylsl dpymrgrmsd  
 61 epsysppvd当地 ggvvmvgtvs rvvesnhpd当地 qsgdwvlgys gwqdydissg ddlvklgdhp  
 121 qnpswsllgvl gmpgftaymg lldigqpkeg etlvvaaatg pvgatvgqig klkgcrvvvg

**FIG. 10I**

181 aggaekcrha tevlgfdvcl dhhaddfaeq lakacpkgid iyyenvggkv fdavlpllnnt  
 241 saripvcglv ssynatelp pfdrlplllma tvlkkirrlq gfiaqdygh rihefqremg  
 301 qwvkedkihy reeitdglen apqtfigllk gknfgkvvir vagdd

**gor (NP 418475)****Nucleotide sequence (SEQ ID NO:121)**

1 atggcaaacac gaattgaatt tcacaaggcac ggtggcccg aagtacttca agccgttagag  
 61 ttcactcctg ccgatccggc ggagaatgaa atccaggtcg aaaataaagc catccgcac  
 121 aattttatcg acacatatat ccgcagcggc cttaaccgc cgccatcgct acccagcgg  
 181 tttaggcaccg aagcagcagg catcgtagt aaagtccgca gtggtgtaaa gcatattaag  
 241 gcaggcgatc gtgtatctc tgccgactcg gcgtttaggcg cttacagctc tggcataac  
 301 attattgcgg aattaaagccg gattctgcgtt cggcgaattt cttttgagca agctgcggca  
 361 tccttcctgaa aaggcttaac gggttattat ctgctgcgcgaa aacccatgaa attaaacccc  
 421 gatgaacagt tcctgttcca cgcagcggct ggcggcggtt gcttaattgc ctgcccagtgg  
 481 gcaaaagccc tgggcgcgaa acttatccgc accgttaggaa ccgcgcggaaa agcgcagagc  
 541 ggcctaaag cgggcgcgtt gcaggttattt aactatcgta aagaggatct ggtcgagcgg  
 601 ttaaaagaga tcaccggcgtaa gaaagtg cgcgtgggtt acgatccgt gggcagagac  
 661 acctggaaac ggtcgcttgc ttgcctgcgaa cgcgcggct taatggtagt ttttggcaac  
 721 tcatcagggtt cggttaccgg tggtaactta ggcattctca atcaaaaagg ctcgtttagt  
 781 gtgacacgccttccctgca aggctatatac accacgcggg aggaattaac cgaggccagt  
 841 aatgaactgt tctctttgtat tgccagcgggt gtgattaagg tcgatgtcgc cgagcagcag  
 901 aataatccgc tgaaggatgc gcagcgtgcg catgagattc tggaaagccg ggcgcacgcaa  
 961 gttccagcc tgcgtattcc ataa

**Amino acid sequence (SEQ ID NO:122)**

1 matriefhkh ggpevlqave fptadpaene iqvenkaigi nfidtyirsg lypppslpsg  
 61 lgteaagivs kvgsgvkhh agdrvvyaqas algayssvhni iadkaailp aaisfeqaaa  
 121 sflkgltvyy llrktyeikp deqflhaha ggvigliacqw akalgaklig tvgtaqkaqs  
 181 alkagawqvi nyreedlver lkeitggkv rvvydsvgrd twersldclq rrglmvsfqn  
 241 ssgavtvgnl gilnqkgsly vtrpslqgyi ttreetleas nelfsliasg vikvdvaeqq  
 301 kyplkdaqra heilesratq gssllip

**frmA (NP 414890)****Nucleotide sequence (SEQ ID NO:123)**

1 tcagtaacga attacggttc gaatggattt gccttcatgc atcaggtcga aggccgtcatt  
 61 aatttcatcc aggtctatgg tatgcgtgac aaacgggtcc agatcgatata caccttcat  
 121 cgcacatcc accatgcccc gtaactggaa acgaccttcc acgcgcgcac accgcggacc  
 181 tttccatcagc cgaccagtga ccaactggaa tgacgggtt gagatttctt gacccgcac  
 241 cgcgcaccccg atgatcaccg actgacccca gccgcgggtc gcactttca gcgcgcac  
 301 catcacgtt acgttaccga tgcattcaaa ggtatggtcg atacccatt ttttgatatt  
 361 caacaggaca tcttttacgc gtttgcgtt gtcattccggg ttaatgcgtt cggtagcacc  
 421 gaagcgcacgc gccagatcga atttttcggg gttgggtatcg atagcgataa tccgacccgc  
 481 tttccctgca cgcgcgcgc gaaaccatgc cagaccaatc gcgcacccgc caaacacccgc  
 541 aacaaatca cctggctgaa cttttagctgt gttgtgttacc gcgcacccgc cgggtgtcac  
 601 gccacagccc agcaggcaga cgtgttcatg gtttgcgtt ggattttt ttgcacccgc  
 661 cacttccgcg actacgggtt atttactgaa tttagagcacc cccatgtatc gataaagccg  
 721 ctggccgtt gtaaaaaac ggggtgttcc gtttgcgtt accgccttac cctgggttcc  
 781 gcaaccgcg acacagaggat tagtttgcg agaacgcac agaacgcac aactcacact cgcgcac  
 841 cgcgggtttaa agcgggatca catggcgcg aggtttgcgc ctgggttacc cttcaccgc  
 901 ttcaaccaca acgccccccc cttcgatcc gagaaccacc gggaaatcac cttccgggc  
 961 atgcggag agggtaaatg cgtcgatcg gcaaacgcgg gttatgggtt gtttaccgg  
 1021 cacttccatc ttttgcgtt gtttgcgtt accgcacgc gtttaccgg  
 1081 agcaaatca acggcagcac gtttaccgg

**Amino acid sequence (SEQ ID NO:124)**

1 mksraavafa pgkpleivei dvappkkgev likvthtgvc htadaftlsgd dpegvfpvvl  
 61 ghegavvvve vgegvtvvpd gdhvplyta ecgecefcrs gktnlcavr etqgkglmpd

**FIG. 10J**

121 gttrfsyngq plyhymgcst fseytvvaev slakinpean hehvcllgcg vttgigavhn  
 181 takvqpgdsv avfqlgaigl avvqgarqak agriiaidtn pkkfdlarrf gatdcinpnd  
 241 ydkpikdvll dinkwgidht fecignvnm raalesahrg wqgsviigva vagqeistrp  
 301 fqlvtrgrvwk gsafggvkgr sqlpgrmveda mkgdidlepf vtthmsldei ndafdlmheg  
 361 ksirtviry

**ybdR (NP 415141)****Nucleotide sequence (SEQ ID NO:125)**

1 atgaaaagcat tgacttatca cggcccacat cacgttcagg tagaaaatgt tcccgcattcg  
 61 ggcgttgaac aggcatatcgtatttctg cgttacgg caacggcgat ctgtggctct  
 121 gacccatc tttatcgagg caaaatacct caggttaaac atggcgatat ttttggtcat  
 181 gaatttatgg gggaaatgt tgaaaccgga aaggacgtaa aaaatttgc aaaaaggcgac  
 241 cgagtggtaa ttccgttctgtt cattgttgcgtt ggcgactgtt ttttctgtcg attgtcaacaa  
 301 tatgccgcct gcgaaaatac caatgcgggt aaaggcgctg cgctcaataa aaaacagata  
 361 ccagctccag cggcattgtt tggttatagt cacctgtatg gcccgcgttcc tgggtggcag  
 421 gccaatatg tccgcgtccc taaagggat gtggggccgt taaaagtacc gccttgctt  
 481 tcagatgata aagcgctttt cctttctgtt attctgcaca cggcatggca ggcagcaaaa  
 541 aatgcgcaga tccaacaagg ttcaagcggtt gcagtctatg gtgcgttgc tgcgtggattg  
 601 ttgacaatcg cctgtgcacg gttgctcgtt gccgaacaga tttttgtt tgatcatcat  
 661 ccctaccgcgt tgcatatcgcc cggcaccgc tacggcgcga tcccgcattaa ttttgcattaa  
 721 gacagcgatc cggcacatc aattattgaa caaacggcag gtcaccgggg cgtggatgca  
 781 gtaatagacg cccgtcggtt tgaagcgaaa ggcagccacca cggaaacggcgt gctgactaac  
 841 ctgaaactgg agggcagcag cggtaagcg ttgcgtcgtt gtttgcggc ggtcaggcgt  
 901 ggccgcattt ttagcgtacc gggcgtctac gctggattta ttacgcgtt cctgtttggc  
 961 gacgcctttt ataaagggtt gtcgtttaat atgggacaga cccacgttca cgcacggctg  
 1021 ggagaattat taccgttaat tgagaaaggg ttactgaaac cagaagaaat tgttaccac  
 1081 tatatgcgtt ttaagagggc cggccgggggatgatgatggattt tggaaaacgc tgaagaggag  
 1141 tgccgtaaagg tgattctgtt acccgtgcacaaacgcag aggccgcgcga gaaggccgtt  
 1201 tcaggtctgg tgaatgcgtt gcccgggggaa acaatatgaa

**Amino acid sequence (SEQ ID NO:126)**

1 mkalhyghp hvqvenvpdp gveqaddiil ritataicgs dlhlyrgkip qvhgdifgh  
 61 efmgevvetg kdvknlqkdg rvvipfviac gdccfcrlqq yaacentnag kgaalnkqi  
 121 papaaalfgys hlyggvgggg aeyvrpvkgn vgpfkvppll sddkalflsd ilptawqaak  
 181 naqiqgqssv avygagpvg1 ltiacarllg aeqifvvvdhh pyrlhfaadr ygaipinfde  
 241 dsdpasiiie qtaghrgvda vidavgfeak gtttctvltl lklegssgka lrqciaavrr  
 301 ggivsvpgvya afihgflfg dafdkglfsk mgqthvhawl gellpliekllkpeevth  
 361 ympfeeaaarg yeifekreee crkvilvpgq qsaeeaqkav sglvnampgg ti

**YggP (YP 026187)****Nucleotide sequence (SEQ ID NO:127)**

1 atgaaaacca aagttgtgc tatttatggc aagcgggatg tccgtctgcg cgtatttgc  
 61 ctgcacatcataa tgaattactg gtgagtgtaa tttctgacag cgtctgttta  
 121 tcgacccgtt aacgcggcgat actcgatgtt gaacataac gcgtacccga cgtatttagaa  
 181 aatcatccgg tcattaccgg gcatgaatgt gcccgggtt tttgtcaagt gggtaaaaat  
 241 ctcactggca aatataaaaaa agggcagcgt tttgttgcg aaccggcgat ggggttacca  
 301 agccgatatt cagccggcta cagtcacaa tttttggcgc gcaatgccac ttatatgatt  
 361 attccggaaa tcgcattaa tttttggcgtt gtttgcgtt atcaccgtt atcaccgtt tttttgtt  
 421 gcccgcgtcgc tggcagagcc tatgtgtgc attatttgcg tttatcatgc caattatcac  
 481 accacgcataat atgtttatga gcatgcgtt ggcgtcaaaac ctggccggcaa tattgcactg  
 541 ctggcggtgtt caggtccgtt gggcattggc gctatcgatt acgcattaa cggccggata  
 601 caaccgtcgc ggggtttt ggtcgatatac gacgacaaac gtcgtgcgc ggtacagaag  
 661 ctgtgtccgg tggacttggc ggcgttataa ggcatttgc gttgtatgtt gataccaaa  
 721 gggatgagcg atccctgttca gatgtgtgc ggcgttgcac gatgtccgg gtcgtatgt  
 781 attttttttt atgcggcggtt gctgtgttgc gtttgcgtt gatgtatgtt actggccggaa  
 841 gatggcggtt tgaacttcc tggccggccg acggataaaa acttcaaaat gccgtttaat  
 901 ttctacaacgc tccattacaa cagcacgcac gtcgtccgtt catctggcggtt tcaacccggac

**FIG. 10K**

961 gacatgaaag aggcgattgc ccttagcgcc actgggcagt tacagccgtc gtttatggtg  
 1021 acccatatcg gtggcttggta tgccgtgcca gaaaccgtgc tcaatctgcc ggatatccct  
 1081 ggcgtaaaaa aactcattta taacggcggt accatgccgc tcactgccc tggccatttt  
 1141 gccgaaaaaa gcaaaaccga tccgctgttt aaagagttgg cgccgtgg tgaggaaacg  
 1201 cacggcatct ggaatgaaca ggccgagaaa tatctgctgg cacaatttgg cgttgatatac  
 1261 ggggaggccg cgcaatga

**Amino acid sequence (SEQ ID NO:128)**

1 mktvaaayg krdvrlrvfe lpeitdnell vsvisdsvcl stwkaallgs ehkrvpddle  
 61 nhpvitghec agvivevgkn ltgkykkgqr fvlqpamglp sgysagysye yfggnatymi  
 121 ipeiainlgc vlpvhgsyfa aaslaepmcc iigayhanyh ttqyvyeerm gvkggnial  
 181 lacagpmgig aidyainggi qpsrvvvvdi ddkrlaqvqk llpvelaask gielvyvnkt  
 241 gmsdpvqmlr altgdagfdd ifvyaavpav vemadellae dgclnffagp tdknfvvpfn  
 301 fynvhynsth vvgtsggstd dmkeaialsa tgqlqpsfmv thigldavp etvlnlpdip  
 361 ggkkliyngv tmpltaiadf aekgktdplf kelarlveet hgiwneqaek yllaqfgvd  
 421 geaaq

**YiaY (YP 026233)****Nucleotide sequence (SEQ ID NO:129)**

1 atggcagctt caacgttctt tattcatttc gtgaatgtca tcggcgctga ttcattgact  
 61 gatgcatttga atatgtatggc agattatggta tttacccgtt ccttaattgtt cactgacaat  
 121 atgttaacgaa aatttaggtat ggccggcgat gtgcggaaag cacttggaaaga acgcaatatt  
 181 tttagcgttta tttatgtatgg caccacaccc aacccccacca cggaaaacgt cgccgcaggt  
 241 ttgaaatttac ttaaagagaa taatttgcgtt acgtgtatct cctttaggcgg tggcttctcca  
 301 cacgactgcg caaaaggat tgcgtggc gcagccaatgc gcggcgatata tcgcgattac  
 361 gaaggcgttg accgctctgc aaaaccgcag ctggcgatgt tcgcctatcaa taccacggcg  
 421 ggtacggccct ctgaaatgac ccgtttctgc atcatcactg acgaagcgcg tcatatcaaa  
 481 atggcattttt ttgataaaaca tgcgtactccg ctgcgtttctgc tcaatgactc ctctctgtat  
 541 atttgtatgc cgaagtcaact gaccgcgcgca acgggtatgg atgccttaac gcacgcata  
 601 gaagcatatgtt tttctatttgc cgccacgcgcg atcaactgacg cttgtgcact gaaagccgtg  
 661 accatgatttgcgaaacccat gccgttagcc gttgaagatg gcagtaatgc gaaagcgcgt  
 721 gaagcatatgtt cttatgcgc gtttctcgcc ggttatgggt tcaatatgc ttctctgggt  
 781 tatgttcatgc cgtatggcgca ccagctggc gtttcttcaactgcgc acgttgcaca cggtgtatgt  
 841 aacgcgtttt tgctgcgcgca cgttgcggta ttcaacagca aagtgcgcgc tgcacgtctg  
 901 cgtgactgttgcgctgcgtt gggcgtaac gtgacaggtaaaaacgcgc ggaagggtgt  
 961 gaagcctgcataa acgttgcgtt gcaactgcgca gcaaggaaag tggatatccc ggcaggccta  
 1021 cgcgcactgtt acgttgcgtt agaagatttgc ggggtatggc cgactaatgc cctgaaagat  
 1081 ggcgtggctt tactaaccgcatcc gatccaggca actcacgaag aaattgtggc gatttatcgc  
 1141 gcagcgatgtt aa

**Amino acid sequence (SEQ ID NO:130)**

1 maastffips vnvigadslt damnmmadlyg fttrtlivtdn mltklgmagd vqkaleerni  
 61 fsviydgtqp npttenvaag lkllkenncd svislggspn hdcakgialv aanggdirdy  
 121 egvdrssakpq lpmiaintta gtasemtrfc iitdearhik maivdkhvtp llsvndssl  
 181 igmpksltaa tgmtdalhai eayvsiaatp itdacalkav tmiaenlpla vedgsnakar  
 241 eamayaqfla gmafnnaslg yvhamahqlg gfynlphgvc navllphqv fneskvaarl  
 301 rdcaaamgvn vtgkndaega eacinairel akkvdipagl rdlnvkeedf avlatnalkd  
 361 acgftnpiqa theeivaiyr aam

**FucO (NP 417279)****(Nucleotide sequence (SEQ ID NO:131))**

1 atgatggctt acagaatgtat tcttgcgtt acggcatgtt ttgggtgggg tgctgttggg  
 61 gcttttaaccgtt atgaggttgcgaa acggcggtt tatcagaagg cgctgtatgtt caccgataaaa  
 121 acgctgggtgc aatgcgcgtt ggtggcgaaa gtgaccgata agatggatgc tcgcaggctg  
 181 gcatggcgatcc tttacgcgtt cgtatgttgc aacccaaacaa ttactgttgcg taaagaagg  
 241 ctcgggtgtat tccggatgtt cggcgccgtt tacctgtatgc ctattgggttgg tggcttctcca  
 301 caggataactt gtaaaggcgat tggcattatc agcaacaacc cggagtttgc cgatgtgcgt

FIG. 10L

361 agccttggaaag ggctttcccc gaccaataaaa cccagtgtac cgattctggc aattcctacc  
421 acagcaggta ctgcggcaga agtgaccatt aactacgtga tcactgacga agagaaaacgg  
481 cgcagaaggta tttgcgttga tccgcgtat atcccgcagg tggcgtttat tgacgctgac  
541 atgatggatg gtatgcctcc agcgtgaaa gtcgcgacgg gtgtcgatgc gctcaactcat  
601 gctatttgggg ggtatattac ccgtggcgcg tgggcgttaa ccgatgcact gcacattaaa  
661 gcgatttggaaa tcattgcgtt ggcgcgtgcga gatcggttg ctggtgataa ggtatccggaa  
721 gaagaaatgg cgctcggcga gtatgttgcg gtatgggc tctcgatgt tgggttaggg  
781 ttgggtcgatc gtatggcga tccactggc gcgtttttata acactccaca cggtgttgcg  
841 aacccatcc tggtaaccga tgtcatcggt tataacgtctg actttaccgg tgagaagtac  
901 cgcgatatcg cgcgcgtt gggcgtaaa gtggaaaggta tgagccgttga agaggcgcgt  
961 aatgccctgt ttgaaggcggt gtttgccttc aaccgtatgtc tggttatcc gccacatttgc  
1021 cgtatgttgc gtgtacgaa ggaagacatt cccggactgg cgcaggcggc actggatgt  
1081 gtttgcgttgc gtggcaaccc gcgtgaagca acgttgcagg atattgttgc gctttaccat  
1141 accqcccttgtt aa

Amino acid sequence (SEQ ID NO:132)

1 mmanrmilne tawfgrgavg altdevkrrg yqkalivtdk tlvcqcvvak vtdkmndaagl  
61 awaiydgvvp nptitvvkeg lgvfnqnsad yliaigggsp qdtckaigii snnpefadvr  
121 sleglsptnk psvpilaipt tagtaaeavti nyvitdeekr rkfvcdphd ipqvafidat  
181 mmdgmpalk aatgvdalh aiegyitrga waltdalhik aieiiagral gsvagdkdag  
241 eemalgqyva gmgsfnvglg lvhgmahplg afyntphgvna naillphvmr ynadftgeky  
301 rdiarvngvk vegmsleear naaveavfal nrdrvippnl rdvgvrkedi palaqaaldd  
361 vctggnprea tledivelyh taw

EutG (NP 416948)

Nucleotide sequence (SEQ ID NO:133)

1 atgcaaaaatg aattgcagac cgcgcctttt caggcgttc ataccctgaa tctgcaacgg  
61 gtaaaaacat ttagcggtcc accgggtacg ctttgcggtc cgggctcggt gagcagtgc  
121 ggacagcaag cgcaaacgcg tgggctgaaa catctgttc tgatggcaga cagcttttg  
181 catcaggcag ggatgaccgc cgggctgacg cgtagcctga ccgttaaagg tatcgccatg  
241 acgctctggc catgtccggt gggcgaaccg tgcattaccg acgtgtgtgc accgtggcg  
301 cagttgcgtg agtcaggctg tgatgggtg atcgcgtttg gcggcggctc ggtgtggat  
361 gcggcggaaag ccgtgacgtt gctggtgacg aaccggata gcacgtggc agagatgtca  
421 gaaaccagcgc ttctgcaacc ggcgttgcgg ctgattgcca ttcaactac cccggaaacc  
481 ggctctgaaa ccaccaatgt aacggtgatt atcgacgcgg tgagcggcg caacgagggt  
541 ttagcccatg cctcgctgat gccggatgtg gcgatcctcg acgcccattt gaccgaagg  
601 gtgcgtcgc atgtcacggc gatgaccggc attgtatgcgt taaccatgc cattgaagca  
661 tacagcgccc tgaacgctac accgtttacc gacagtctgg cgattgtgc cattgcgtatg  
721 attggcaatcg cgtgtccgaa aegcggtggc tacggtcacg accttgcgc ggcgagagc  
781 atgttgcgtgg cttcatgtat ggcggaaatg gcgttttcca gtgcgggtct tgggttgtgc  
841 cacgcgtatgg cgcatcagcc ggccgcggcg ctgcataattt cgcacggctc cgcgaacgcc  
901 atgttgcgtgc caacggtgtat ggaatttaac cggatggttt ctgcgtgaac ctttagtcag  
961 attggtcggg cactgcaac taaaaaatcc gacgatcggt acgctattaa cgcggtaagt  
1021 gagctgattt cgaaagtgg gattggtaaa cgactggcg atgttggtgc gacatctgc  
1081 cattacggcg catgggcgcg ggccgcgtg gaagatattt gtctgcgcag taaccgcgt  
1141 accggccagcc tggagcagat tgtccggctg tacgcagcgg cgcaataa

Amino acid sequence (SEQ ID NO:134)

1 mqnqlqtalf qafdtlnlqr vktfsvvpt lcgpgsvssc gqqaqtrglk hlfvmadsfl  
61 hqagmtaglt rsltvkgiam tlwpcpvgep citdvcaava qlresgcdgv iafggsvld  
121 aakavtllvt npdstlaems etsvlqprlp liaipttagt gsettnvtvi idavsgrkqv  
181 lahaslmpdv aildaalteg vpshvtamtg idalthaiea ysalnatpft dslaigaiam  
241 igkslpkavg yghdlaares mllascmagn afssaglgc hamahqpgaa lhiphglana  
301 mllptvmefn rmvcrerfsq igralrtkks ddrdainavs eliaevgigk rlgdvgtsa  
361 hygawaqaal ediclrsnpr tasleqivgl yaaaq

**FIG. 10M****YqhD (NP\_417484)****Nucleotide sequence (SEQ ID NO:135)**

```

1 atgaacaact ttaatctgca caccctaacc cgattctgt ttggtaaagg cgcaatcgct
61 ggtttacgcg aacaaattcc tcacgatgt cgcgtattga ttacctacgg cggccgcagc
121 gtaaaaaaaa ccggcggttct cgatcaagtt ctggatgccc taaaaggcat ggacgtgctg
181 gaatttggcg gtattgagcc aaaccggct tatgaaacgc tgatgaacgc cgtgaaactg
241 gttcgcgaac agaaagtgac tttccgtct gcgggttggcg gcggttctgt actggacggc
301 accaaattta tcgccccagc ggctaactat ccggaaaaata tgcattccgt gcacattctg
361 caaacggcg gtaaagagat taaaagcgc accccgatgg gctgtgtgt gacgctgcca
421 gcaaccgggtt cagaatccaa cgcaggcgcg tgatctccc taaaaccac aggacacaag
481 cagcggttcc attctgccc ttgcagccg gtatggccg tgctcgatcc ggttataacc
541 tacaccctgc cgccgcgtca ggtggtaac ggcgtatgtgg acgccttgtt acacaccgtg
601 gaacagtatg ttaccaaacc ggttgatgcc aaaatttcagg accgtttcgc agaaggcatt
661 ttgctgacgc taatcgaaga tggtcggaaa gccctgaaag agccagaaaa ctacgatgtg
721 cgcgccaacg tcatgtggc ggcgacttag gcgcgtgaacg gttgattgg cgctggcgt
781 cccaggact gggcaacgca tatgtctggc cacgaaactga ctgcgtatgc cggtctggat
841 cacgcgcaaa cactggctat cgtccgtct gcactgtgga atgaaaaacg cgataccaaag
901 cgcgctaagc tgctgcaata tgctgaaacgc gtctgaaaca tcactgaaagg ttccgatgat
961 gagcgtattt acgcccgcgt tgccgcaacc cgaatttct ttgagcaatt aggctgccc
1021 acccacctct ccgactacgg tctggacgcg agtccatcc cggcttgct gaaaaaaactg
1081 gaagagcagc gcatgacca actggcgaa aatcatgaca ttacgatgg tgcagccgc
1141 cgtatatacg aagccgcccc cttaa

```

**Amino acid sequence (SEQ ID NO:136)**

```

1 mnnfnlhpt rilfgkgaia glreqiphda rvlitygggs vkkvqvlldqv ldalkgmdvl
61 efggiepnpa yetlmnavkl vreqkvtfll avgggvsvldg tkfiaaaany penidpwhil
121 qtggkeiksa ipmgcvltlp atgsesnaga visrkttgdk qafhsahvqp vfavlvpvyt
181 ytpprprqvan gvvvdafvhvt eqyvtpkpvd kqdrrfaegi lltliedgpk alkepenydv
241 ranvmwaatq alnligagv pqdwhamhlg heltamhgl haqtlaivlp alwnekrdtk
301 rakllqyaer vwnitegsdd eridaaaat rnffeqlgpv thlsdyglg ssipallkkl
361 eehgmtqlge nhditldvsr riyeaar

```

**AdhE (NP\_415757)****Nucleotide sequence (SEQ ID NO:137)**

```

1 atggctgtta ctaatgtcgc tgaacttaac gcactcgtag agcgtgtaaa aaaagccccag
61 cgtgaatatg ccagttcac tcaagagcaa gttagacaaaaa tcttccgcgc cggccgtctg
121 gctgctgcag atgctcgat cccactcgcg aaaatggccg ttgccaatc cggcatgggt
181 atcgtcgaag ataaagtgtat caaaaaccac tttgcttctg aatatatcta caacgcctat
241 aaagatgaaa aaacctgtgg ttgtctgtct gaagacgaca cttttggtagt catcaactatc
301 gctgaaccaa tcggtatatt ttgggtatc gttccgacca ctaacccgac ttcaactgtct
361 atctttcaaat cgctgatcag tctgaagacc cgtacacgcca ttatcttc cccgcaccccg
421 cgtgcaaaag atgcccaccaa caaacgggtt gatatcggtc tgcaggctgc tatecgctgcc
481 ggtgcgttca aagatctgtat cggctggatc gatcaacccctt ctgttgaact gtataacgca
541 ctgatgcacc acccagacat caacctgtatc ctcgcactg gtggccggg catggtaaa
601 gccgcataca gctccggtaa accagctatc ggtgttagggc cgggcaacac tccagttgtt
661 atcgtgaaa ctgctgatata caaacgtgca ttgcattctg tactgtatc caaaacccctt
721 gacaacggcg taatctgtc ttctgaacag tctgtgttt ttgttactc tgtttatgac
781 gctgtacgt aacgtttgc aacccacggc ggctatctgt tgcagggtaa agagctgaaa
841 gctgttcagg atgttaccc gaaaaacgggt ggcgtgaacg cggctatctgt tggcagccaa
901 gcctataaaaa ttgctgaaact ggcaggctt tctgtaccag aaaacaccaa gatttgcata
961 ggtgaagtga ccgttggta taaaaggcga cccgtcgac atgaaaaact gtccccggact
1021 ctggcaatgt accgcgttaa agatttcgaa gacgcggtag aaaaagcaga gaaactgggtt
1081 gctatggcg gttatcggtca tacccatctgc ctgtacactg accaggataa ccaaccggct
1141 cgcgtttctt acttcgggtca gaaaatgaaa acggcgcgtt ccgttattaa caccggcg
1201 tctcagggtt gttatcggtca cctgtatcc ttccaaactcg caccctccct gactctgggt
1261 tgggtttctt ggggtggtaa ctccatctct gaaaacgttg gtccgaaaca cctgtatcaac

```

**FIG. 10N**

1321 aagaaaaccc ttgctaagcg agctgaaaac atgttggc acaaacttcc gaaatctatc  
 1381 tacttccgccc gtggctccct gccaatcgcg ctggatgaag tgattactga tggccacaaa  
 1441 cgtgcgtca tcgtgaactga ccgttccctg ttcaacaatg gttatgtca tcagatcact  
 1501 tccgtactga aagcagcagg cggtgaaact gaagtcttct tcgaagttaga agcggaccgg  
 1561 accctgagca tcgttgcgtaa aggtgcagaa ctggcaaaact ccttccaaacc agacgtgatt  
 1621 atcgcgtggc gtgggtgttc cccatggac gccgcgaaga tcatgtgggt tatgtacgaa  
 1681 catccggaaa ctcacttcga agagctggcg ctgcgttta tggatatccg taaacgtatc  
 1741 tacaagttcc cgaaaatggg cggtgaaagcg aaaatgatcg ctgtcaccac cacttctggt  
 1801 acagggtctg aagtcaactcc gtttgcgggt gtaactgacg acgtactgg tcagaaatata  
 1861 cccgctggcag actatgcgt gactccggat atggcgattt tcgacgcca cctgggttatg  
 1921 gacatgccga agtccctgtg tgcttcgggt ggtctggacg cagaactca cgccatggaa  
 1981 gcttatgttt ctgtactggc atctgagttc tctgtatggc aggctctgca ggcactgaaa  
 2041 ctgctgaaag aatatctgcc aggtcctac cacgaagggt ctaaaaatcc ggtagcgcgt  
 2101 gaacgtgttc acagtgcgcg gactatcgcg ggtatcggt ttgcgaacgc cttcctgggt  
 2161 gtatgtcaactatggcga caaaactgggt tcccagttcc atattccgca cggctctggca  
 2221 aacggccctgc tgatttgtaa cggttattcgtc tacaatgcga acgacaaccc gaccaagcag  
 2281 actgcattca gccagttatga ccgtccgcag gtcgcgcgtc gttatgtca aattgccgac  
 2341 cacttgggtc tgagcgcacc gggcgaccgt actgtctgtc agatcgagaa actgtctggca  
 2401 tggctggaaa cgctgaaagc tgaactgggt attccgaaat ctatccgtca agctggcggt  
 2461 caggaagcag acttcctggc gaacgtggat aaactgtctg aagatgcatt cgatgaccag  
 2521 tgcacccggc ctaaccggcg ttacccgctg atctccgacg taaaacagat tctgtctggat  
 2581 acctactacg gtcgtgatta tggatggat gaaaactgcag cgaagaaaaga agctgctccg  
 2641 gctaaagctg agaaaaaagc gaaaaaatcc gcttaa

**Amino acid sequence (SEQ ID NO:138)**

1 mavtnvaeln alvervkkaq reyasftqeq vdkifraaai aaadaripla kmavaesgm  
 61 ivedkvikh faseniyinay kdekctgvls eddtfgtiti aepigicgi vpttnptsta  
 121 ifkslislkt rnaiifspk rakdatnkaa divlqaiaa gapkdligwi dqpsevlsna  
 181 lmhpfdinli latggpgmvk aayssgkpa gvgagnpvc idetadikra vasvlmskrf  
 241 dngvicaseq svvvvdsvyd avrerpahg gyllqqkelk avqdvilkn alnaaivgqp  
 301 aykiaelagf svpentrkili gevttvdese pfaheklsp lamyrakdfe davekaeklv  
 361 amggightsc lytdqdnqpa rvsyfqqkmk tarilintpa sqggigdlyn fklapsltlg  
 421 cgswggnsis envgpkhlin kktvakraen mlwhkpkpsi yfrrgs1pia ldevitdghk  
 481 ralivtdrfl fnngyadqit svlkaagvet evffeveadp tlsivrkgae lansfkpdvi  
 541 ialgggspmd aakimwvmye hpethfeela lrfmdirkri ykfpkmvgvka kmiavttts  
 601 tgsevtpfav vtddatggky pladyaltpd maivdanlv mmpks1caf gldavthame  
 661 ayvsvlasef sdgqalqalk llkeylpasy hegsknppvar ervhsaattia gianfanaflg  
 721 vchsmahklg sqfhiphqla nallcnvir ynandnptkq tafsqydrpq arryaeiad  
 781 hlglsapgdr taakieklla wletlkaelg ipksireavg qeadflanvd klsedafddq  
 841 ctganprypl iselkqilld tyygrdyveg etaakkeaa akaekkakk a

**dkgB (NP 414743)****Nucleotide sequence (SEQ ID NO:139)**

1 atggctatcc ctgcattttgg tttaggtact ttccgtctga aagacgacgt tggatattca  
 61 tctgtataa cggcgcttca acttgttat cgcgcattt ataccgcaca aatctatgt  
 121 aacgaagccg cagtaggtca ggcgatttca gaaaatggcg tgccacgtca tgaactctac  
 181 atcaccacta aatctggat tggatattctc agcaaaagaca aattgtatccc aagtctgaaa  
 241 gagagcctgc aaaaatttgcg taccgattat gttatctca cgctaatcca ctggccgtca  
 301 cccaaacgtg aagtctctgt tgaagatgtt atgcaggcgc tgctggaaagc caaaaaaca  
 361 gggctgacgc gtggatgtt tatttccac ttacacgtcc cgttgatggaa aaaaagcgatt  
 421 gctgctgttgc tgctgaaaaa catcgctact aaccagattt aactctctcc ttatctgca  
 481 aacccgtaaag tgggtgcctg ggctaaacag cacggcatcc atattacttc ctatatgac  
 541 ctggcgtatg gtaaggccct gaaagatgag gttattgtct gtatcgcagc taaaacacaat  
 601 gcgactccgg cacaagtgtat tctggcgtgg gctatggggg aagggtactc agtaattcc  
 661 tcttctacta aacgtaaaaa cctgaaaatg aatcttaagg cacaatattt acagcttgc  
 721 gccgaagata aaaaagcgat cgccgcactg gattgcaacg accgcctgg tagccccggaa  
 781 ggtctggctc ctgaatggga ttaa

**FIG. 10O****Amino acid sequence (SEQ ID NO:140)**

1 maipafglgt frlkddvvis svitalelgly raidtaqiyd neaavgqaia esgvprhely  
 61 ittkiwienl skdklipslk eslqklrtdy vdltlihwps pndevsveef mqalleakqk  
 121 gltreigisn ftiplmekai aavgaeniat nqielspylq nrkvawakq hgihitsymt  
 181 laygkalkde viariaakhn atpaqvilaw amgegysvip sstkrknles nlkaqnqlqd  
 241 aedkkaiaal dcndrlvspe glapewd

**YdjG (NP 416285)****Nucleotide sequence (SEQ ID NO:141)**

1 ataaaaaaaga tacctttagg cacaacggat attacgcattt cgcgaaatggg gttggggaca  
 61 tgggccatttggcgggggtcc tgcatggaat ggccatctcg atccggaaat atgtattgtat  
 121 acgattcttg aagcccatcg ttgtggcatt aatctgattt atactgcgcc aggatataac  
 181 tttggcaata gtgaaggattt cgtccgttcg gggttaaaaaa aactgcggcc tgacacagggtt  
 241 gtagtagaaaa ccaaatgcgg cattgtctgg gaacgaaaag gaagtttattt caacaaagtt  
 301 ggcgatcggc agttgtataa aaacccccc ccggaaatcta tccgcgaaga ggtacgcacgc  
 361 agcttgcac cgtctgggtt tgattacatc gatatctaca tgacgcactg qcagtcgggt  
 421 ccggccatttt ttacgcgcgat cgctggaaact gtcgcagtgc ttaatgagtt aaagtctgaa  
 481 gggaaaatttgcgcgtataagg cgctgcataac gtcgcgtctg accatatccg cgagtatctg  
 541 caatatgggt aactggatatttatttcggcgg aaatacagta tcctcgaccg ggcaatggaa  
 601 aacgaactgc tgccactatg tcgtgataat ggcattgtgg ttcagggttta ttccccggcata  
 661 gaggaggat tgggacccgg caccatcaact cgtgattacg ttccgggggg cgctcggggca  
 721 aataaaagtct ggttccacgc tgaaaacatg ctgaaagtga ttgatatgct tgaacagtgg  
 781 cagccactttt gtcgtcgta tcagtgcaca attcccaactc tggcactggc gtggatattta  
 841 aaacagagtg atttaatctc catttttagt ggggctactg caccggaaaca ggtacgcgaa  
 901 aatgtcgcgg cactgaatat caacttatcg gatgcagacg caacattgtat gagggaaatgg  
 961 gcagaggccc tggagcgta a

**Amino acid sequence (SEQ ID NO:142)**

1 mkkiplgtd itlsrmglgt waigggpawn gdldrqicid tileahrcgi nlidtapgyn  
 61 fgnsevivvgg alkklpreqv vvetkgivw erkgslnkv gdrqlyknls pesireevaa  
 121 slqrldiydi diymthwqsv ppfftpiaet vavlnekkse gkiraigaan vdadhireyl  
 181 qygeldiiqa kysildrame nellplcrdn givvvqvyspl eqglltgtit rdyvpggara  
 241 nkvwfqrenm lkvidmleqw qplcaryqct iplalawil kqsdlisils gatapeqvre  
 301 nvaalnlnls dadatlmrem aealer

**YeaE (NP 416295)****Nucleotide sequence (SEQ ID NO:143)**

1 atgcaacaaa aaatgattca atttagtggc gatgtctcac tgccagccgt agggcaggga  
 61 acatggata tgggcgaaga tgccagtcag cgccaaacacg aagttgtctgc actacgcgcg  
 121 ggcattgaac tcggtttaac cctcattgtat accgcggaaa tgtatggcga tggcggtgcc  
 181 gaaaagggtgg ttggggaaacg attaacccgtt ctgcgagaga aggtctttct cgtctctaaa  
 241 gtcctatccgtt ggaatgtcgcc cgggcaaaaa gcgataaaatg catgcgaaggc cagtttacgc  
 301 cgtctcaata ctgattatct cgtatcccttac ttattacact ggtctggcag ttgcgtttt  
 361 gaagagactg tcgcagcgat gggaaaatttgc gatgcggccagg gggaaaatccg ccgcgtggggc  
 421 gtttctaacc ttgattatgc tgatgtcag gaaactctggc agtgcgggg gggaaaatcag  
 481 tgtgcacta atcagggtgt ttaccatctc gttcacgag gaattgagta cgatctactc  
 541 ccctgggtgcc agcaacacgca gatgccgggtg atggcttaca gtcgtttagc ccaggccggg  
 601 cggttgcgcgca atggactgtt aaaaaacgcgc gtagtcaacg aaattgcaca tgctcacaat  
 661 atcagcgcgg cacaagtattt gttggcgtgg gtgatcgtc atcagggtgt gatggcgtt  
 721 cccaaagcgg ccacgatgc ccatgtccaa caaaatgcgg ctgtgcgttga ggtcgaactt  
 781 tcttcagcgg aattagctat gctggataag gcatatccgg caccaaaagg aaaaactgcg  
 841 ctggatatgg tgtga

## FIG. 10P

**Amino acid sequence (SEQ ID NO:144)**

```

1 mqqkmiqfsg dvslpvavggg twymgedasq rktevaalra gielgltlid taemyadgga
61 ekvvgealtg lrekvflvsk vypwnaggqk ainaceaslr rlntdyldly llhwsgsfaf
121 eetvaamekl iaqgkirrwg vsnlidyadmq elwqlpggnq catnqvlyhl gsrgieydl
181 pwccqqqqmpv maysplaqag rlrngllkna vvneiahahn isaaqvllaw vishqgvmai
241 pkaatiahvq qnaavlevel ssaelamldk aypapkgkta ldmv

```

**dkgA (NP 417485)****Nucleotide sequence (SEQ ID NO:145)**

```

1 atggctaatac caaccgttat taagctacag gatggcaatg tcatgccca gctgggactg
61 ggcgtctggc aaqcaagtaa tgaggaagta atcaccqcca ttcaaaaagc gttagaagtg
121 ggttatacgct cgattgatac cgcgcggcc tacaagaacg aagaagggtgt cggcaaagcc
181 ctgaaaaatg cctcagtc当地 cagagaagaa ctgttcatca ccactaagct gtggAACGAC
241 gaccacaagc gcccccgca agccctgctc gacagcctga aaaaactcca gcttgattat
301 atcgacctct acttaatgca ctggcccgtt cccgctatcg accattatgt cgaagcatgg
361 aaaggcatga tcgaattgca aaaagaggga ttaatcaaaa gcatcggcgt gtgcaacttc
421 cagatccatc acctgcaacg cctgattgtat gaaactggcg tgacgcctgt gataaaccag
481 atcgaacttc atccgctgtat gcaacaacgc cagctacacg cctggaacgc gacacacaaa
541 atccagaccc aatccctggag cccattagcg caaggaggga aaggcggtt cgatcagaaa
601 gtcattcgcg atctggcaga taaatcagcg aaaaacccgg cgccatgtat tatccgctgg
661 catctggata gccgccttgtt ggtgatcccg aaatcggcata caccttcacg tattggcggaa
721 aactttgtat tctgggattt ccgtctcgac aaagacgaac tcggcgaat tgcaaaaactc
781 gatcaggcga agcgtctcgg tcccgatcct gaccagttcg gcggctaa

```

**Amino acid sequence (SEQ ID NO:146)**

```

1 manptviklq dgnvmpqlgl gvwqasneev itaiqkalev gyrsidtaaa ykneeqvgka
61 lknasvnree lfittklwnd dhkrpreall dslkkqlqldy idlylmhpvp paidhyveaw
121 kgmielqkeg liksigvcnf qihhlqlrlid etgvtpvinq ielhplmqqr qlhawnathk
181 iqteswspla qggkgvfdqk virdladkyg ktpaqivirw hldsglvvip ksvtpsriae
241 nfdvwdfrld kdelgeiakl dqgkrlgpdp dqfgg

```

**yajo (NP 414953)****Nucleotide sequence (SEQ ID NO:147)**

```

1 atgcaatacaca accccttagg aaaaaccgac cttcgcgtt cccgactttt cctcggtgt
61 atgacctttg gcgagccaga tcgcggtaat cacgcatacactgca cactgccga agaaaggcagc
121 cgtccctataa ttaaacgtgc acttgcggc ggcataaaatt tctttgatac cgccaaacagt
181 tattctgacg gcagcagcga agagatcgactt ggtcgccac tgcgggattt cgcccgctgt
241 gaagacgtgg tcgttgcgc当地 caaatgttc catcgcgtt gtgatttacc ggaaggattt
301 tcccgatcgc当地 aaatttgc当地 ctctatcgac gacagcctgc gacgtctcgg catggattat
361 gtcgatatacc tcgaaattca tcgcgtggat tacaacacgc cgatcgaaga gacgctggaa
421 gcccctcaacg acgtggtaaa agccggaaa ggcgttata tcggcgc当地 atcaatgcac
481 gtcgtcgactt ttgc当地 cacttgc当地 caaaaacagc acggctggc gcaatgttgc
541 agtatgcagg atcactacaa tctgtatcgat cgtgaagaag agcgc当地 gacgtaccactg
601 tggatcagg agggcgtggc ggttatacc tcggagccgc tggcaagggg ccgtctgacg
661 cgtccgtggg gagaaactac cgacacactg gtgtctgtat aggtggggaa aaatctctat
721 aaagaaagcg atgaaaatga cgc当地 agatc gcaagacgatc taacaggcgt cagtgaagaa
781 ctggggggc当地 caccgacaca agttgcgtg ccgttgc当地 tgatcacc gggcattgcc
841 gcaccgatta tcggaacttc ggc当地 agaa cagcttgc当地 agtattgaa cgc当地 ggat
901 atcactttga agccggaaaca gattgc当地 ctggaaacgc cgtataaaacc gcatcctgtc
961 gtaggattta aataa

```

**Amino acid sequence (SEQ ID NO:148)**

```

1 mqynplgktd lrvsrlclgc mtfgepdrgn hawtlpeess rpiikraleq qinfddtans
61 ysdgsseeeiv gralrdfarr edvvvatkvf hrvqdlpeg sraqilrsid dslrrlgmdy
121 vdilqihrd yntpieetle alndvvkagk aryigassmh asqfaqalel qkqhgwqfv
181 smqdhynliy reeerempl cyqegvavip wsplargrlt rpwgettarl vsdevgknly

```

FIG. 10Q

241 kesdendaqi aerltgvsee lgatraqval awllskpgia apiigtsree qldellnavd  
301 itlkpegiae letpykphv vgfk

YghZ (NP\_417474)

Nucleotide sequence (SEQ ID NO:149)

```
1 atggctcggt tagcgaatcc cgaacgttac gggcagatgc aataccgcta ttgcggaaaa  
61 agtggtttac gcctgccgc gttatcgctc gtttatggc acaatttcgg tcacgttaac  
121 gcgcgtgaat cacagcgtc gatcctcggt aaagcgttt atttggcat tacgcacttt  
181 gattttagcca acaattacgg gccgcctcca ggaagcgcag aagagaactt tggcgcctg  
241 ctgcgggagg atttgcgcg ttatcgat gaactgatta tctctaccaa ggctggctac  
301 gatatgtggc ccggccctta cggtctggc gtgtcacgt aatacctgct cgccagcctc  
361 gaccaaagcc tgaagcgat ggggtttag tatgtcgata tctttactc tcatcgctc  
421 gatgaaaata cgccgatgga agaaaccgcc tctgcgctgg ctcatcggt acaaagcggt  
481 aaggcgctgt atgtcggtat ctcctcttac tcgcccagagc ggacgcaaaa aatggtcgag  
541 ttgcgtcgcg agtggaaaat tccgtgtta attcatcaac ctgcgtacaa ttactgaac  
601 cgctgggtgg ataaaagcgg cctgtggat accctgcaaa ataacggcgt ggcgtgtatt  
661 gccttactc ctctggctca gggattgtcg accggaaaaat atctcaacgg cattccgcaa  
721 gattcacgga tgcatcgta agggataaaa gttcggtgc tgacacgg aatgttacc  
781 gaagccaacc tcaacagcct gcgcatttgc aatggaaatgg cacagcagcg tggacaatca  
841 atggcgcaaa tggcgtaag ctgggtgtc aaagatgatc gcgtgcgtc ggtattgatt  
901 ggtgccagcc gcgcggagca actagaggag aacgtgcagg cgctgaataa tctgacatctt  
961 agcaccaagg agctggcgca gattgatcag catatcgccc atggcgagct gaatctgtgg  
1021 caggcgcttt ccgataaaatg a
```

Amino acid sequence (SEQ ID NO:150)

```
1 mvwlancery gqmqyrycgk sglrlpalsl glwhnfhvn alesqrailr kafdlgithf  
61 dlannyygppp qsaeenfgrrl lredfaayrd elistkagy dmwpypgysg gsrkyllasl  
121 dqslkrmggle yvdifyshrv dentpmeta salahavqsg kalyvgissyspertqkmve  
181 llrewkipll ihqpsynlln rwvdksqlld tlqnngvqci aftplaqgll tgkyllngipq  
241 dsrmhregnk vrgltpkmlt eanlnslrll nemaaqrgqs maqmalswll kddrvtsvli  
301 gasraeqleev nvgalnnltf stkelaqidq hiadgelnlw qassdk
```

Tas (NP 417311)

(Nucleotide sequence (SEQ ID NO:151)

1 atgcaatatac accgtatacc ccacagttcg ctggaaagtca gcacgcgtgg gcttggcacf  
61 atgacgtttg gtgaacagaa cagcgaagcc gacgcggc acacaactcga ctagccgtc  
121 gtcaggcga ttaaccctat cgacgttgcc gaaatgtacc cagtacctcc ggcggccgaa  
181 acgcaagggt taaccgaaac ctacgtcgcc aactggctgg cggaaacatgg cagccgcgaa  
241 aagttaatta tcgcctccaa agtgagcgga cctgtcgccgata ataatgacaa gggcatccgc  
301 ccggatcagg cgctggatcg gaagaatatc cgcgaagcgc tgcatgacag cctcaagcgc  
361 ctacagactg attacctcga tctttatcag gtgcactggc cgcagcggcc gaccaactgc  
421 ttcggcaaac tcggttatag ctggacggat tctgcgcctg cggtttcgct gctggatacg  
481 ctggacgcac tggcagagta ccaacgcgcg gaaaaattc gttataatcgg cgtgtcgaac  
541 gaaaactgcac ttggcgtaat ggcgtacctg catctggcgcg acaaaccacga tctgcgcggt  
601 attgtcacca ttccagaaccc ttacagtctg ttaaaccgcga gttttgaagt aggtctggca  
661 gaagtccgcg agtatgaagg ggtcgaactg ctggccattt cgtgcctggg ttccggcacf  
721 ctgaccggga aatatctcaa ttgtccaaaa cccgcgtggc cacgtataac gctctttagt  
781 cgggtccaccg gctatagcgg tgagccaaacg caaaaagccg tcgcggcgtatgttgc  
841 gccagacgtc atggcctgga ccctgtcgat atggcgcctcg cggtaaccgc  
901 ttttgtgcct gcaactctgt gggcgcacacc acgatggatc agctgaaaac taacatcgaa  
961 agtttgcattc tggagttaa cgaagacgtat ttagctgaaa ttgaaggcgtt gcatcagggtt  
1021 tatacttatac cggcaccata a

Amino acid sequence (SEQ\_ID\_NO:152)

1 mqyhriphss levstlglgt mtfgceqnsea dahaqldyav aqginlidva emypvpprpe  
61 tqgltetyvg nwlakhgsre kliiaskvsg psrnndkgir pdqaldrkni realhdsllkr  
121 latdyldlyq vhwpqrptmc fgklqyswtd sapavslldt ldalaeyqra qkiryiqvsn

**FIG. 10R**

181 etafgvmryl hladkhdlpr ivtiqnpysl lnrsefvgl evsqyegvel laysclgfgt  
 241 ltgkylnak pagarntlfs rftrysgeqt qkavaayvdi arrhgldpaq malafvrrqp  
 301 fvastllgat tmdqlktnie slhlelsedv laeieavhqv ytypap

**YdhF (YP\_025305)****Nucleotide sequence (SEQ ID NO:153)**

1 atggttcagc gtattactat tgcgcccaa gccccggagt tttcccgaaa ttgtatgggc  
 61 tactggcgat tgatggactg gaatatgtcc gcccggccagc tggtcagttt tattgaagag  
 121 catctggatc tcggcgtgac caccgtggac catgctgata tttatgggtt ctatcgtgc  
 181 gaagcggcgt ttggcgaggc actgaaactg gcaccctcacc tgcgtgaacg gatggaaatc  
 241 gtcagtaaat gcggtatcgc gacgaccgcg cgtgaagaaa acgtcattgg tcattacatc  
 301 actgaccgcg atcacatcat taagagcgcgaa acacagtcgc taattaatct cgccaccgat  
 361 catctggatt tgctgttaat ccaccgacca gaccggttaa tggatgccc tgaagtggcg  
 421 gacgcgttca aacatctgca tcagagcggc aaagtgcgtt atttggcgat atcgaacttt  
 481 acgcctgcgc aatttgcctt gttgcaatca cgtctggcgat ttacccttgc cactaatcag  
 541 gtggaaatat ccccggtgca tcagccgtta ctgctggat gcacgcgtcg ccaactacaa  
 601 caactgcgtg ttctgtccat ggcgtggcc tgccttgggtt gtggtcgtct gtttaatgtat  
 661 gattatttc agccgtgcg tgatgaaactg gctgtgggtt cagaggagtt aaacgcgggc  
 721 tcgattgaac aggtgggtt cgcctgggtt ttacgtttac catcgccagcc gctgccaatt  
 781 atcggttcag gtaaaattga gcgctacgg gcaagctgtcg aagcagaaac actgaaaatg  
 841 acccgtaac aatggttcg tatccgtaaa goggcactgg ggtacgacgt accgtaa

**Amino acid sequence (SEQ ID NO:154)**

1 mvqriniapq gpefsrfvng ywrlmdwnms arqlvsfiee hldlgvttd hadiyggyqc  
 61 eaafgealkl aphrlrermei vskcgiatta reenvighyi tdrdhiiksa eqslinlatd  
 121 hldlllihrp dplmdadeva dafkhlhqsg kvrhfgvsnf tpaqfallqs rlpftlatnq  
 181 veispvhqpl lldgtldqlq qlrvrmpaws clgggrlfnd dyfqplrdel avvaeelnag  
 241 sieqvvyawv lrlpsqlpli igsdkiervr aaveaetlkm trqqwfirk aalgydvp

**YdbC (NP\_415924)****Nucleotide sequence (SEQ ID NO:155)**

1 atgacgacgca atacatttac tctcggtaca aaatccgtta accgtcttgg ttatggcgcg  
 61 atgcaactgg caggcctggc agttttggc ccccccacggat atgcacgtt cgctataacc  
 121 gtgtcgctg aggccgtggc attgggggtc aatcatatttgc ataccagcga cttttatgg  
 181 cccgacgtca ccaatcagat tatcccgaa ggcgtttatc cttactctgac cgacctgaca  
 241 attgtcacta aaatttggcgc gccggcgtggc gaggacgcattt cctgggttgc cgcattttct  
 301 cccgcacggc tgcaaaaaggc ggtgcacgtt aatctacgtt atctcggtt ggacgtgcgt  
 361 gatgtggta acctgcgcgt tatgtatggg gatgggtatc gcccagcggc aggtatcgatt  
 421 gaggccagcc tgaccgtgtt ggcagagatg caacaacaag gcctggtaaa acatatttggc  
 481 ctgaccaacg tcacaccgcac gcagggttgc gaggcgcgc acatatttggc aattgtctgt  
 541 gtgcaaaaacg aataacaacat cgcgcacccgt gctgtatgtt caatgatttgc tgcttggcc  
 601 cacgatggca ttgcctacgtt ggcgttcttc cccgtcgggg gctttacacc gctgcaatcg  
 661 tccacactt ccgatgttgc tgcgagcctt ggttgcacac caatgcagttt ggcgtggcg  
 721 tggctgttac agcgttaccat gaatatttttgc tgcgttccat ggcacgttgc ggttgcgc  
 781 ttacgggaga atatggctgc tgaaaatttgc catctttgtt aggaagtgtt gtctacgtt  
 841 gatgttattt cgcgagaata a

**Amino acid sequence (SEQ ID NO:156)**

1 mssntftlgt ksvnrlgyga mqlagpgvfg pprdrhvait vlrealalgv nhidtsdfyg  
 61 phvtinqiire alypysddlt ivtkigarr edaswlpafs paellqkavhd nlrnlglvdv  
 121 dvvnrlrvmmg dghgpaegsi easltvlaem qqqlvkhig lsnvptqva earkiaeivc  
 181 vqneyniahr addamidala hdgiayvpff plggftplqs stlsdvaasl gatpmqvala  
 241 wllqrsplnil lipgtssvah lrenmaekl hlseevlsts1 dgisre

FIG. 10S

ybb0 (NP 415026)

Nucleotide sequence (SEQ ID NO:157)

1 atgactcata aagcaacgga gatccctgaca gttaaaggta tgcaaaaatc ggtcttaatt  
61 accggatgtt ccagtggaat tggcctggaa agcgcgctcg aattaaaacg ccagggttt  
121 catgtgctgg caggttgcgg gaaaccggat gatgttgagc gcatgaacag catgggattt  
181 accggcgtgt tgatcgatct ggattcacca gaaagtgttg atcgcgcgac cgacgagggtg  
241 atcgccctga ccgataaattg tctgtatggg atctttaaca atgcggatt cggcatgtat  
301 gccccccctt ccaccatcag ccgtgcgcag atggaacacg agtttccgc caacttttc  
361 ggcgcacacc agctcaccat gcgcctgtt cccgcgtatg taccgcacgg tgaaggggcgt  
421 atttgtatgta catcatcggt gatggggatta atctccacgc cgggtcggt cgcttacqcg  
481 gccagtaataat atgcgtgtt ggcgtgttca gatgcactgc gcatggatg ggcgcacacg  
541 ggaataaaag tcagectgtat cgaaccgggtt cccattcgta ctgcgttac cgcacaacgtc  
601 aaccagacgc aaagtgtataa accagtgcgaa aatcccccgcg tcgcccggc cttagcgttgc  
661 hgaccggaaag cgggtgttgcg caaagtagcgc catgttttta tttagcgagaa gcccgaagatg  
721 cgctatccgg tgacgtgtt gacctggcg gtaatggtgc ttaagcgctt gctgggggg  
781 cqcggtatgg acaaaaatatt qcagggtqa

Amino acid sequence (SEQ\_ID\_NO:158)

1 mthkateilt gkvmqksvli tgcssiggle salelkrqgf hvlagcrkpd dvermnsmgf  
61 tgvlidldsp esvdraadev ialtdnclyg ifnnagfgmy gplstisraq meqqfsanff  
121 gahqltmrll pamphgegr ivmtssvmlg istpgrgaya askyaleaws dalrmelrhs  
181 qikvsliepg pirtrftdnv ngtqsdkpve npgiaarftl gpeavvdkvr hafisekpkm  
241 rypvtlvتوا vmvvlkrllpg rvmdkilgg

yohF (NP 416641)

Nucleotide sequence (SEQ ID NO:159)

1 atggcacagg ttgcgattat taccgcctcc gatcgaaaaa tcggcaaaaga gtgcgcgtta  
61 ttactggcgc agcaggggtt tgatatttgtt attacctggc actcagatga agaagggcga  
121 aaagataccg cgcgtgagggt agtttagccac ggcgtacgtg cggagatcgt gcagctggat  
181 ctccggcaatc taccagaagg ggcactggcg ctggagaaaac tcattcaacg gctggggcgc  
241 attgtatgtgc lggllaaalaa lgcggglyca atgaccaaag cgccgtttct tgatatggct  
301 tttgtatgagt ggcgcgaagat ttttacgtt gatgtcgatg gtgcattttt atgctcgcaa  
361 attgcggcgtc gtcagatgggt gaaacaaggg cagggcggtc gcatcatcaa cattacgtcg  
421 gtacatgaac atacgcgcgt gccggatgcc agcgcctaca cagccgctaa acatgcgtc  
481 ggtgggttaa ccaaagcgat ggcgcgtggag ctggtcaggc ataagatttt ggtgaacgca  
541 gtgcgcgcctg gggcgatcgca cacgccaatg aatggcatgg atgacagcga cgtgaagccc  
601 gacgcggagc cttcgattcc cttgcggcgt tttggcgcaa cgcattcgat tgccacgcctg  
661 gtgggtgtggc ttgttcggaa gggcgcaaat tacaccaccc ggcagtcgtt gatagtggat  
721 ggcggcttta tgttggcgaa tccacagttt aaccggaaat ag

Amino acid sequence (SEQ\_ID\_NO:160)

1 maqvaiitas dsgigkecal llaqqgfdig itwhsdeega kdtarevvsh gvraeivqlq  
61 lgnlpegala lekliqrngr idvlvnna mtkapfldma fdewrkiftv dvdgaflcsg  
121 iaarqmvkqg qggriinitv vhehtplpda saytaakhal ggltkamale lvrhkilvna  
181 vapgaiatpm ngmddsdvvp daepsiplrr fgatheiiasl vvwlcsegan yttggslivd  
241 ggfmlanpqf npe

YciK (NP\_415787)

Nucleotide sequence (SEQ ID NO:161)

```
1 atgcattacc agccaaaaca agatttactc aatgatcgca ttatcttgtt gacggggacc  
61 agcgatggta ttggcgtgta agccgcgtat acgtatgcac gctatggtc gacagtgttt  
121 ctgttgggcc gtaatgaaga aaaattacgt caggttagcca gccacataaa cgaagaaact  
181 gggcgtcagc cacagtgggtt tattctcgat ttgtgtaccc gcacgtccga aaattgcca  
241 caactggcac agcgcattgc cgtaattat ccgcgtctgg atggtgtttt gcataatgc  
301 ggattgtctcg gcgatgtttt cccaatggac gaacaaaatc cgcagggtctg gcaggacgtc  
361 atgcaggatca acgttaatgc cacctttatq ctcacccaaqq cactgtttcc tttattactc
```

**FIG. 10T**

421 aaatcgacg ccggttcaact ggtctttact tcatcaagcg ttggacgtca gggacgagcc  
 481 aactgggtg catatgcagc gtcgaaattt gccaccgaag ggatgtatgc ggtactggcc  
 541 gatgaatatac agcagcgcct gcgtgtcaac tgcatthaacc caggcggtag cgccaccgc  
 601 atgcgtgcca ggcgcctccc gaccgaagat ccacagaaac ttaaaacacc cgctgatatac  
 661 atgccgctct acctctggct gatggcgat gacagccgc gtaaaaccgg catgacctt  
 721 gacgccccaaac cgggcccgtaa accaggaatt tcccaatga

**Amino acid sequence (SEQ ID NO:162)**

1 mhyqpkqdll ndriilvtga sdgigreaaam tyarygatvi llgrneeklr qvashineet  
 61 grqpqwfild lltctsenqc qlaqriavny prldgvlnha gllgdvcpmis eqnpqvwqdv  
 121 mqvnvnatfm ltqallplll ksdagslvft sssvgrqqra nwgayaaskf ategmmqvla  
 181 deyqqrlrvn cinpggtrta mrasafpted pqkllktpadi mplylwimgd dsrrktgmtf  
 241 daqpgrkpgi sq

**Ygff (NP 417378)****Nucleotide sequence (SEQ ID NO:163)**

1 atggctatacg cacttgtac tggtggcagt cgccgcacatcg ggcgggcaac tgcattactg  
 61 ttggcgcaag aagggtatacg ggtggcggtt aattatcagc aaaacctcca cgccggcgca  
 121 gaagtgtatga acttaataac gcaagccggt ggcaaagcat tcgtgttcca ggcggatatac  
 181 agcgacgaaa accaggtcgt tgcgtatgtt acagcaatcg atcagcacga tgaaccgcta  
 241 gcagcgctgg tcaataaacgc cgggatcttg tttaccatgc gacccgttga aacaccttacc  
 301 gcagagcgaa tcaaccgagt accttccacc aacgtgacgg gatattttct ctgctgcccgc  
 361 gaggcgttaa aacgcattgcgctt aacatgcgatgttgc gccggctat cgtcaatgtc  
 421 tcttcgggtt cctcacgggtt gggttcgcca gggaaatatgttgc ggcattacgc ggcattcgaaa  
 481 ggggcgttgc atacgttaac caccggacta tcgctggaaatgc tgcggcgca ggggatccgc  
 541 gttaactgcg tgcggccagg gtttatatttgc accgaaatgc acgcacggc cggcggccct  
 601 ggacgcgttc atcgcgttaa gtcgaacatc cccatgcgac gttgggaca ggcagaagag  
 661 gtcgcgcagg ccattgtctg gctactaagt gataaaggct citacgtcac gggaaagttt  
 721 atcgattttgg cggggggaa ataa

**Amino acid sequence (SEQ ID NO:164)**

1 maialvtggs rgigratall laqegytvav nyqqnlhaaq evmnlitqag gkafvlqadi  
 61 sdenvvvamf taidqhdepl aalvnagil ftqctvenlt aerinrvlst nvtgyflccr  
 121 eavkrmlkn ggsggaiavnv ssvasrlgsp geyvdyaask gaidlittgl slevaaggir  
 181 vncvprgfiy temhasggep grvdrvksni pmqrqgqaee vaqaiwvils dkasyvtgsf  
 241 idlagk

**YghA (NP 417476)****Nucleotide sequence (SEQ ID NO:165)**

1 atgtctcatt taaaagaccc gaccacgcg tattacactg gtgaatatcc caaacagaaaa  
 61 caaccgacgc caggcatcca ggcgaagatgc acaccggtagt cggattgcgg cgagaaaaacc  
 121 tatgttggta gcggcgccct gaaagatcgt aaagcactgg tgacaggggg cgattccgga  
 181 ataggtcgcg ctgcgcctat cgcttacgcg cgtgaagggg ctgacgtggc gatcagttat  
 241 cttcccggtt aagaagaaga cgctcaggat gtggaaaaga tcattgaaga atgcggacgc  
 301 aaagccgttc tgctgccagg cgatttaagc gatgagaaat ttgcccgttc gctgggttac  
 361 gaagcgcaca aggcgtttagg cgggctggat attatggcgc tggtcgcccgg gaaacaggtt  
 421 gccattccgg atattgcaga cctcaccacgc gaacagtttc aaaagacatt tgccattaac  
 481 gtttccggc tggctgttgc aaccaggaa gcgatcccc tgctaccgaa aggtgcgaatgt  
 541 atcatcacca cttcgtaat ccagcatac cagccaaatgc cgcatttactt ggactatgcg  
 601 gctacgaagg cggcgatttc gaactacagc cgtggcttgg caaaacaggtt cgcggagaaaa  
 661 ggtattccgg tgaatattgt cgcgcacggc cccatctggc cagcactgca aatttccggc  
 721 ggacaaacgc aggataagat ccccgatgtt ggtcggcaaa cgcggatgaa acgtgcgggg  
 781 caacccggccg aactggcccc tggatgttgc tatctggca gtcaggagtc gagtaacgtc  
 841 accgcagaag tgcacccgtt gtcggccggc gaggatattag gttaa

**Amino acid sequence (SEQ ID NO:166)**

1 mshlkdpptq yytgeypkqk qptpgiqakm tpvpdcgekt yvgsgrlkdr kalvtggdsg

**FIG. 10U**

61 igraaaaiaya regadvaisy lpveeedaqd vkkiiieecgr kavllpgdls dekfarslvh  
 121 eahkalggld imalvagkqv aipdiadlts eqfqkftfain vfalfwltqe aipllpkgas  
 181 iittssiqay qpsphllyda atkaailnys rglakqvaek girvnivapg piwtalqisg  
 241 gqtqdikpqf qqqtppmkrqf qpaclapvyv ylasqessyyv taevhgvcgg ehlg

**YjgI (NP 418670)****Nucleotide sequence (SEQ ID NO:167)**

1 atgggcgttt ttacaggtaa gacagttctc atccctcggtg gcagtcgtgg tatacggtgcc  
 61 gctatcgtac gtcgttgcgt caccgatggg gccaatgtac gattcaccta tgcgggggtcg  
 121 aaagatgccc ctaaacgcct ggacacaagag actggagcga cagcagtatt cacagatagt  
 181 gctgacagag acgctgtcat tgatgtcggtt cgtaagagcg gcgcattgga tatacggtgg  
 241 gttaaatgcag gtattggcgat ctttggcgag gcccgttggaaataatgcga cgatattgtat  
 301 cgcctttca aaatcaatat tcatgtcct tatcatgcct ctgttgaaagc cgccccggcaq  
 361 atgcccgaag gcggggcgat cttaatcattc ggctcggtga atggcgtatcg tatgcgttgc  
 421 gcaggcatgg ctgcttatgc cggcagcaaa tctgcccgtc aaggcatggc gcgcgggctg  
 481 gcccgtgatt ttggaccccg tgggatcacc attaacgtcg tccagccagg gccaatttgat  
 541 accgacgcta atccccccaa cggggccaaatcg cgcgatatgt tgcatagttt gatggctatc  
 601 aaaagacatg ggcaaccggaa agaggctcgat ggtatggctc catgggttagc agggccagaa  
 661 gccagttttt ttaccggcgc gatgcataacc attgtatggcg cgtttggcgc ataaa

**Amino acid sequence (SEQ ID NO:168)**

1 mgafgtktvl ilggssrgiiga aivrrfvtdg anvrftyags kdaakrlaqe tgatavftds  
 61 adravidvv rksgaldilv vnagigfvge alelnaddid rlfkinihap yhasveaarg  
 121 mpegrilii gsvngdrmpv agmaayaask salqqmargl ardfgprgit invvqpgpid  
 181 tdanpangpm rdmlhslmai krhggpeeva gmawlagpe asfvtgamht idgafga

**YdfG (NP 416057)****Nucleotide sequence (SEQ ID NO:169)**

1 atgatcggtt tagtaactgg agcaacggca gttttgggtg aatgcattac tcgtcggtt  
 61 attcaacaag ggcataaaatg tatcgccact ggcgtcgcc aggaacggtt gcaggagtt  
 121 aaagacgaaac tgggagatata tctgtatatac gcccaactgg acgttcgca ccgcgcgc  
 181 attgaagaga tgctggcatc gcttcctgcg gagtgggtgca atattgtat cctgttaaat  
 241 aatgcggcc tggcggttggg catggggct ggcataaaag ccagcggtga agactggaa  
 301 acgtatggt atacaacaa caaaggcctg gtatatacg cgcgcgcgt cttaccgg  
 361 atgttgaac gtaatcatgg tcatattt aacattggct caacggcagg tagctggcc  
 421 tatgcgggtg gtaacgttta cggtgccacg aaacgcgttgc ttgcgtcaat tagcctgaat  
 481 ctgcgtacgg atctgcatttacgg taccgggtg cgcgtcaccg acatcgaaacc gggctgg  
 541 ggttgtaccg agtttccaa tgtccgttt aaaggcgatg acgtaaagc agaaaaaacc  
 601 tataaaaata ccgttgcatt gacggcagaa gatgtcagcg aagccgtctg gtgggtgtca  
 661 acgtgcctg ctcacgtcaa tatcaatacc ctggaaatga tgccgttac ccaaagctat  
 721 gccggactga atgtccaccc tcagtaa

**Amino acid sequence (SEQ ID NO:170)**

1 mivlvtgata gfgecitrrf iqqqghkviat grrqerlqel kdelgdnllyi aqlrvnrar  
 61 ieemlaslpa ewcnidilvn naglalgmep ahkasvedwe tmidtnnkgl vymtravlp  
 121 mvernhghii nigstagswp yaggnyvgat kafvrqfsln lrtdlhgtav rvtdiepglv  
 181 ggttefsnvrk kgddgkaekt yqntvaltpe dvseavwwvs tlpahvnint lemmptqsy  
 241 aglnvhraq

**YgcW (NP 417254)****Nucleotide sequence (SEQ ID NO:171)**

1 atgtcaatcg aatctctcaa tgcgttctca atggatttt tctccctgaa agttaaaacc  
 61 gcaattgtta cgggtggaa tagcgggtta ggccaggcat ttgcgttgc gttggccaaa  
 121 gctggcgcaatatactttat tccttagtttgc tcaagata acggcggaaac aaaggaaatg  
 181 attgaaaaac agggtgttgc ggtggacttc atgcagggtgg gtatcaccgc agaaggcgc

**FIG. 10V**

241 ccgcagaaga ttatcgctgc ttgctgttag cgtttcggta cagttgatat tctggtaac  
 301 aatgccgta ttttaagct gaataaggcg ctggacttcg gtcgtgccga ctggatccg  
 361 atgattgatg tgaacctgac cgccgcattc gagttaaagct atgaagctgc aaaaattatg  
 421 atcccccaga aaagccgcaaa aattattaat atctgttcat tttctctta cttaggtgaa  
 481 caatggtcac ctgcataattc tgccactaaa catgtcttg ccgggttacaa caaagcttat  
 541 tgtgatgaac taggtcaata taatatttcg gtaaatggta tcgcccctgg ctattatgca  
 601 acagatatta cgctggcgc acgcagtaat ccagaaaccatc acagcgcgt tcttgatcat  
 661 attccggcaaa accgttgggg cgataactcgat gatttaatgg gcgcagccgt attccctcgca  
 721 agtccggcat cgaattatgt caacgggcat ttattagtgg ttgatggcgg ttattnatgt  
 781 cgctaa

**Amino acid sequence (SEQ ID NO:172)**

1 msieslnafs mdffslkgkt aivtggnsrl qqafamalak aganifipsf vkdngetkem  
 61 iekqgvvedf mqvgvitaega pqkiaacce rfgtvdlvn nagiicklnkv ldfgradwdp  
 121 midvnlttaaf elsyeakim ipqksgkiin icslfsylgg qwspsaysatk halagftkay  
 181 cdelgqyniq vngiapgyya tditlatrsn petnqrvlldh ipanrwgdtq dlmgaavfla  
 241 spasnyvngh llvvdcgylv r

**UcpA (NP\_416921)****Nucleotide sequence (SEQ ID NO:173)**

1 tcagataaccg acgctaaccg tctccggcag tggctgccc ccatcaatca cattctgtgt  
 61 accggtaaaa tagctggatt catccgatgc gaggaaaggcc gccaggatcgc cgacttccag  
 121 cggatcggcg aggccgacgca tcgggattgc tttcgccatt tcagtcagca ccgactctgg  
 181 atcttccggg ttcgacttgc gggcaatgct ttccgcccatt ggtgtgcgc cgtatccgg  
 241 gccaatggcg ttaacgcgaa taccagactg cgccgtactcc accgcacgcg attttgtcag  
 301 gccaacaatc gccgcttgc ttaaggcgta cgccgttgc ccaggatcgg ccacccatata  
 361 accagtactgaaagacatca tcacaaatgcg accatctttt cgggcaatca tctccggcag  
 421 caccgccttc gtgacgttcc atacgcctt aatattgtat tcaatatggaa atacgcgatc  
 481 gtcatcgctc atatcgagga aactgcccag acgcacaaacg cctgcgttat tcaccaggat  
 541 atcaatgcgc ccttcgtcg tttgatagtc gccgcttcc acgcggggc  
 601 acgcacatcg ggcacaaccg ccgtacagcg atgaccacga ccacacaggat cgtccggcag  
 661 ctttcgtatc tcaggggaga tatccagcaa gatttagttc ggcacatgac gtgcaaaatg  
 721 tctggcaatt ctttcgtccaa ttccctgcaaa tgccggcgtt atcagtgtctg tcttgcgg  
 781 gagtttaccc at

**Amino acid sequence (SEQ ID NO:174)**

1 mgkltgkta itgalggige giartfarhg anlilldisp eiekkladelc grghrctavv  
 61 advrpdasva aaikrakeke gridilvnna gvcrlgsfld msdddrdfhi dinikgvwnv  
 121 tkavlpemria rkdgrivmmms svtgdmvadp getayaltka aivgltsla veyaqsgirv  
 181 naicpgyvrt pmaesiargs npedpesvlt emakaipmrr ladplevgel aaflasdess  
 241 yltgtqnvid ggstlpetsv vgi

**EntA (NP\_415128)****Nucleotide sequence (SEQ ID NO:175)**

1 atggatttca gcggtaaaaa tggctgggtt accggcgca gtaaaggat cggctacgcc  
 61 accggcgctgg cgtttgttga ggcgggagcg aaagtttacag gttttgtatca agcgtttact  
 121 caggagcaat atcccttgc gaccgaatgt atggatgttgc cgcacgttgc gcaaggctcg  
 181 caagtgttgc agcgacttgc agtggaaacg gagcgactgg acgcgttgc caatgcggcg  
 241 ggaattttac gcatggcg gaccgttgc ctcagttaaaggactggca gcaactttt  
 301 gccgttaacg tcggcggttgc gtttaaccttgc ttccagcaaa ccatgaacca gtttcgcgt  
 361 cagcggggcg gggcgatgtt cactgtggcg tccgacgcgc cgcacacgc gcttattggc  
 421 atggatgttgc atggcgatc gaaaggcgctt gtttttttttggcgatggatggatggatgg  
 481 gaactggcg gtagccggcg ggcgttgc gtttttttttttttttttttttttttttttttttt  
 541 atgcaacgcgca cgctgtgggtt gaggatgc gccgaagaac agcgatattcg cggctttggc  
 601 gagcagtttta aactcgccat tccgtgggg aaaatcgccc gtccacaa gatcgccaaac  
 661 acgattttgc tcctcgccctc tgaccctgcg agccatatttttccatcagggatattttttt  
 721 gatggcggtt caacgctggg ggcataaa

**FIG. 10W****Amino acid sequence (SEQ ID NO:176)**

1 mdfsgknvwv ttagkgigya talafveaga kvtgfdqafq qeqypfatev mdvadaaqva  
 61 qvcqrlllaet erldalvnaa gilmgatdq lskedwqqtf avnvggafnl fqqtinqfrr  
 121 qrqgaivtva sdaahtrprig msaygaskaa lkslalsvgl elagsvrn vvspgstdtd  
 181 mqrtlwvsdd aeqqrirgfg eqfklgiplg kiarpqeian tilflasdla shitlqidivv  
 241 dggstlga

**FolM (NP 416123)****Nucleotide sequence (SEQ ID NO:177)**

1 atgggtaaaa cccagccctt gccaatatta attactggcg gaggtcgtcg catccgcctc  
 61 gcccgcatttcat taatcaaaaag caaccgtga ttgtcagcta tcggacacac  
 121 tatccagcca ttgtatggact gattaatgca ggtgcgcagt gtatttcaggc tgatttttcg  
 181 accaacgcg qtgtatggc gtttgccat gaagtactaa aaagcaccca tggctgcgt  
 241 gcttatttgc ataacgcacg tgcgtggat gggaaaaac cgggtgcgc actggccac  
 301 gtactggctt gcatgtatgca gatccacgtt aatacccat acctgtctaa ccatgcgt  
 361 gaaagattac tgcgtgggc cggacacgc gccagcgata tcatttcattt taccgattat  
 421 gtgttggagc gggtagcga caaacatatt gcgtatgtc caagcaaagc ggcactggat  
 481 aatatgaccc gctcggttgc ccgcacgctg gcacccgaa tgaaagtga ttctattgcg  
 541 ccatcgctga tcttgttta tgaacatgtat gatgccat atcgacaaca ggcgtctaat  
 601 aaatcaactga tgaaaaccgc gcctggcgag aaagaagtga tcgacctggat cgattactta  
 661 cttaaccgtt gcttgcac cggacgcagt ttccacttg atggcggtcg tcatctgcgt  
 721 taa

**Amino acid sequence (SEQ ID NO:178)**

1 mgktqplpil itgggrriegl alawhfinqk qpvisyrrth ypaidqlina gaqcqadfs  
 61 tndgvmafad evlkstghlr ailhnasawm aekpgaplad vlacmmqihv ntpylnhal  
 121 erllrghha asdiihftdy vvergsdkhi ayaaskaald nmtrsarkl apevkvnbia  
 181 psilfnehd daeyrqqaln kslmktpage kevidlvdyt ltscfvtgrs fpldggrhhr

**HdhA ( NP 416136)****Nucleotide sequence (SEQ ID NO:179)**

1 gtgttaatt ctgacaacct gagactcgac ggaaaatgcg ccatcatcac aggtgcgggt  
 61 qcaggattt gtaaagaaat cggcatttaca ttgcgcacag ctggcgcata tgggtggc  
 121 agtataattt acgcccacgc agttaaccat gtttagacgc aaattcaaca actgggtgg  
 181 caggcattt cctgcgttg ttagtattact tccgaacagg aactctctgc actggcagac  
 241 ttgtatca gtaagctggg taaagtttat attctgttaca aacacccgg tggcggtgg  
 301 cctaaaccgt ttgtatgc aatggccgtt ttgcgcgtt cttatgtactt gaatgtgttt  
 361 tctttttttt atctgtcaca actgttgcg ccagaaatgg aaaaaatgg cgggtggcg  
 421 attctgtatgc tcaacttgcg tggcgccagaa aataaaaaatgg taaacatgc ttcttatgc  
 481 tcatactaaatg ctcggccat tcatactgtc agaaatatgg cgtttgaccc ggtgaaaaaa  
 541 aatattcggg taaatggcat tgcgcgggg gcaatattaa ccgatccctt gaaatccgtt  
 601 attacaccg aaatttgc aaaaatgtt cagcacacgc cgatcagacg tctggccaa  
 661 ccgcaagata ttgctaacgc agcgctgttc cttgtctgc ctgtcgag ctgggtaaac  
 721 ggacaaattt tcaccgttca cgggtgggg gtacaggagc tcaattaa

**Amino acid sequence (SEQ ID NO:180)**

1 mfnlsdnrlrd gkcaittgag agigkeiait fatagaszvvv sdinadaanh vvdeiqqlgg  
 61 qafacrcdit seqelsalad faisklgkvd ilvnnagggg pkpfmpmad frayelnvf  
 121 sffhlsqlva pemekngggv iltitsmaae nkninmmtsya sskaaashlv rnmafndlgek  
 181 nirvngiapg ailtdarksv itpeiekml qhtpirrlgq pqdianaalf lcspaaawvs  
 241 qgiltvsggg vgeln

**HcaB (NP 417036)****Nucleotide sequence (SEQ ID NO:181)**

**FIG. 10X**

1 atgagcgatc tgcataacga gtcattttt attaccggcg gcggatcggtt attagggtctg  
 61 gcgcgttgtcg agcgatttat cgaagaaggc gcgcagggtt ccacgctggaa actgtcgccg  
 121 gcaaaaagtgcg ccagtctgcg tcagcgattt ggcaaacata ttctggcggt ggaaggtaac  
 181 gtgacctgtt atgccgatta tcaacgccc gtcgatcaga tcctgactcg ttccggcaag  
 241 ctggattgtt ttatcgccaa tgccggatc tgggatcaca atgcctcaact ggtaataact  
 301 ccccgagaga cgctcgaaac cggttccac gagctgttta acgtcaatgt tctcggttac  
 361 ctgctggcg caaaagcctg cgctccggcg ttaatcgcca gtgaaggcag catgattttc  
 421 acactgtcaa atgccgctcg gtatcctggc ggccgtggcc cgctgtacac cgccagtaaa  
 481 catgccgcaa ccggacttat tcgccaactg gcttatgaac tggcacccaa agtgccgggtg  
 541 aatggcgtcg gcccgtgtgg tatggccagc gacctgcgcg gcccacagggc gctcgccaa  
 601 agtgaacacct cgataatgca gtctctgacg ccggagaaaa ttgcccacat ttaccgctg  
 661 caattttcc cgcaaccggc ggattttacg gggccgtatg tcatgttgc atcgccggc  
 721 aataatcgcg cattaagcgg tgtgatgatc aacgctgtatc cgggttttagc gattcgccgc  
 781 attcgccacg tagcggctgg gctggatctt taa

**Amino acid sequence (SEQ ID NO:182)**

1 msdlhnnesif itgggsglgl alverfieeg aqvatlelsa akvaslrqrf gehilavegn  
 61 vtcyadyqra vdqiltrsgk ldcfignagi wdhnaslvnt paetletghf elfnvnlgy  
 121 llgakacapa liasegsmif tlsnaawypg gggplytask haatglirql ayelapkvrv  
 181 ngvgpcgmas dlrgpqalqq setsimqslt pekiaailpl qffpqpadft gpyvmltsrr  
 241 nnralsgvimi nadaglairg irhvaagldl

**SrlD (NP 417185)****Nucleotide sequence (SEQ ID NO:183)**

1 atgaatcagg ttgcgttgtt catcggttgtt gggcaaacct taggcgcgtt cctgtgcac  
 61 ggtctggctg ccgggggta tcgcgtcgat gttgtcgata ttcaagagcg caaagccgca  
 121 aatgtggcac aagaaattaa cgccgaatat ggtgaaagta tggcgtacgg tttttgtgt  
 181 gacgccacta gcgagcaaa cgttctggcg ctctctcgat gggtagatga aatctttgt  
 241 cgctgtggatt tgctgtctca cagcgcggc atagcaaaag cagcctttt cagcacttc  
 301 cagctcggcg attttgaccc ttgcgtacag gtatcgttgg tgggttattt cctgtgtgc  
 361 cgtgaatttt cgcgtttat gatccgcgc gggatttcagg ggcgcattat tcagatcaac  
 421 tggaaatccg gcaaaagtggg cagcaaacac aactctggct acagcgcgc gaaattttgt  
 481 ggctcgggc tgactcaatc actggcgctg gatctggcg agtacggcat tacgtgtcat  
 541 tcactgtatc tcggtaacat gctgaaatcc cccatgttcc agtactgtt gccacaataac  
 601 ggcaccaagc tgggtatcaa accggatcaa gtcgagcgtt attacatcga caaagtaccg  
 661 ctcacaaacgcg gctgcgatca tcaagatgtt ctgcattatc tgctgttca cgccagtcc  
 721 aaggcgtcgatc actgcacccgg acagtgcgtt aatgtcaccg cgggtcaggt gatgttctga

**Amino acid sequence (SEQ ID NO:184)**

1 mnqavvvigg gqtlgafch glaaeeyrva vvdijqsdkaa nvaeinaey gesmaygfga  
 61 datseqsvla lsrvdveifg rvdlvlysa iakaafisdf qlgdfdrslq vnlvgyflca  
 121 refsrlmird giqgriiqin sksgkvgskh nsgysaakfg gvgltsal dlaeygitvh  
 181 slmlgnllks pmfqslppqy atklgikpdq veqyyidkvp lkrgcdyqdv lnmllyasp  
 241 kasyctgqsi nvtggqvmf

**KduD (NP 417319)****Nucleotide sequence (SEQ ID NO:185)**

1 atgattttaa gtgcattttc tctcgaaaggtaa agttgcgg tcgtcactgg ttgtgataact  
 61 ggactgggtc agggatggc gttggggctg gcgcaggcg gctgtgacat tttggcatt  
 121 aacatcggtt aaccgactga aaccatcgat caggtcacag cgctggggcg tcgtttttta  
 181 agcctgaccg ccgatctcgat aagattgtt ggtattccag cactgctggatcgcgcgtt  
 241 gcccggatggt gtcattatcgat tttccgtt gataacgcgg gattgttgcggcaggat  
 301 gctctcgatc tcagcgaaaa ggactggcgc gatgtcatcg acctgaaatcg caagacgtt  
 361 ttcttcgtt ctcaggcgc ggcggaaacac tttatcgccg aaggcaatgg cgccaggatt  
 421 atcaatatcg cgtcaatgtt ctcttccat ggcgggatcc gtgtgccttc ttataccgc  
 481 tcaaaaaggcg gctgtatggg tggatggcgatc ttatggcgatc acaatactca acaactacgg  
 541 attaatgtt atgcgatagc cccgggttac atggcgacca acaatactca acaactacgg

**FIG. 10Y**

601 gcagatgaac aacgttagcgc ggaaattctc gaccgcattc cagctggtcg ttggggactg  
 661 ccgagtgacc tcatggggcc gatagtgttc cttgcctcca ggcgttcaga ttatgtgaat  
 721 gtttatacca ttgcgttgg tggcggttgg ctggcgctt aa

**Amino acid sequence (SEQ ID NO:186)**

1 milsafsleg kvavvtgc dt glggmalgl aqagcdi vgi n i v e p t e t i e q v t a l g r r f l  
 61 s l t a d l r k i d g i p a l l d r a v a e f g h i d i l v n n a g l i r r e d a l e f s e k d w d d v m n l n i k s v  
 121 f f m s q a a a k h f i a q g n g g k i i n i a s m l s f q g g i r v p s y t a s k s g v m g v t r l m a n e w a k h n  
 181 i n v n a i a p g y m a t n n t q q l r a d e q r s a e i l d r i p a g r w g l p s d l m g p i v f l a s s a s d y v n  
 241 g y t i a v d g g w l a r

**I dnO (NP 418687)****Nucleotide sequence (SEQ ID NO:187)**

1 atgaacgatc tattttca ct ggcaggaaaa aatatcttga ttaccggttc agcacaggc  
 61 attggcttt tactggca ac cggcctgggta aatataatggc cacaataat tattaatgtat  
 121 attactgccc aacgcgcaga actgtctgt aaaaaactcc accaggaggg tattcaggcc  
 181 gtgtccgcac ct ttttaatgt tactcataaa catgaaaattt atgcccgcgt tgaacatata  
 241 gaaaaggaca tcggccccat tgatgtgctg gtgataaacg ccggtatcca gogccgtcat  
 301 ccttttactg agttccctga acaagagtgg aatgatgtga tcgcagtaaa ccagaccgccc  
 361 gtgttccctgg tatacga aacgcgcaga ggttaactcg tccatgggtt aacgcgcaggc aggtttaaatgg  
 421 attaatattt gtcgcgtca aagcgaactg ggacgtgaca ccatcaccccc ttatgccgca  
 481 tcgaaaagggg cggtaaaaat gtcacccgc ggcacgtgc tcgagctggc gogccacaat  
 541 attcaggta ca cggatttgc gcccggctat ttcaaaaacag aatgactaa agcactgggtt  
 601 gaggacgaag ctttcaccgc ctgttgtgc aaacggaccc cccgcgcacg ctggggagat  
 661 ccgcaggaac tgattgggtc tgccgttgc ct tttcttcaa aagcctctga ttctgtaaac  
 721 ggcacactgt tg ttttgtga tggcgcatg ttatgtggctg tttaa

**Amino acid sequence (SEQ ID NO:188)**

1 mndlfslagk nilitgsaq igfl latglg kyga qiiind itaeraelav ek lhqegiqa  
 61 vaapfnvthk heidaavehi ekdipidvl vnnaqiqrrh pftefpeqew ndviavnqta  
 121 vflvsqavtr hmverkagkv inicsmqsel grdtitpyaa skgavkmtr gmcvelarhn  
 181 iqvnqia pgy fktemtkalv edeaftawlc krtpaarwgd pqeligaavf lsskasdfvn  
 241 ghllfv dggm lvav

**FabG (NP 415611)****Nucleotide sequence (SEQ ID NO:189)**

1 atgaattttg aaggaaaaat cgcactggta accggtgcaa gccgcggaaat tggccgcgc  
 61 attgtctgaaa cgctcgac c cgtggcgcg a a a g t t a t t g g c a c t g c g a c c a g t g a a a a t  
 121 ggcgctcagg cgatcagtga ttat ttaggt gccaacggca a a g g t c t g a t g t g a a t g t g a a g t g  
 181 accgacccgg catctatcg a t c t g t t c t g a a a a a t t c g c g a a t t t g g t g a a g t g  
 241 gatatcctgg tcaataatgc cggtatcact cgtgataacc t g t t a a t g c g a a t g a a a g a t  
 301 gaagagtgg a acgatattat c g a a a c c a a c c t t c a t c t g t t t c c g t c t g t a a a g c g  
 361 gtaatgcgcg ctatgatgaa a a a g c g t c a t g t g c t t a a t c a t c g g t t c t c t g t g g t  
 421 ggtaccatgg gaaatggcgg tcaggccaa tacgctgcgg c g a a a g c g g g c t t g a t c g g c  
 481 ttcatgaaat cactggcgcg c g a a g t g c g t c a c g c g g t a t t a c t g t a a a c g t t g t g c t  
 541 ccgggcttta t g a a a c c g g a c a g a c a c g t g c g t a a a t c g c c a a c g c g g t a t c  
 601 ctggcgcagg t t c c t g c g g g t c g c t c g g c g g a a a t c g c c a a c g c g g t a t c  
 661 t t c c t g g c a t c c g a c g a a g c a g t t a c a t c a c g g g t a a a t c g c c a a c g c g g g  
 721 atgtacatgg tctga

**Amino acid sequence (SEQ ID NO:190)**

1 mnfegkialv tgasrgigna iaetlaarga kvigtatsen g a q a i s d y l g a n g k g l m l n v  
 61 tdpasiesvl ekiraefgev dilvnnagit rdhllmrmdk eewndiietn lssvfrlska  
 121 v m r a m m k k r h g r i i t i g s v v g t m g n g g q a n y a a k a g l i g f s k s l a r e v a s r g i t v n v v a  
 181 pgfieddmtr alsddqrugi laqvpagrlg g a q e i a n a v a f l a s d e a a y i t g e t l h v n g g  
 241 mymv

**FIG. 10Z****FabI (NP 415804)****Nucleotide sequence (SEQ ID NO:191)**

1 atgggttttc ttccggtaa gcgcattctg gtaaccggtg ttgccagcaa actatccatc  
61 gcctacggta tcgctcaggc gatgcaccgc gaaggagctg aactggcatt cacctaccag  
121 aacgacaaaac tgaaaggccg cgtagaagaaa ttgcccgc aattgggttc tgacatcgaa  
181 ctgcagtgcg atgttgcaga agatgccagc atcgacacca tttcgctga actggggaaa  
241 gtttggccga aatttgacgg tttcgatcac tctattgggtt ttgcacctgg cgatcagctg  
301 gatggtgact atgttaacgc cgttacccgt gaaggcttca aaattgcccc caacatcagc  
361 tcctacagct tcggttgcataat ggcaaaaagct tgccgcctca tgctgaatcc gggttctgcc  
421 ctgctgaccc tttcctacact tggcgcttag cgccgtatcc cgaactacaa cgttatgggt  
481 ctggcaaaag cgtctctgga afgcaacgtg cgctatatgg cgaacgcgt gggccggaa  
541 ggtgtcgctg ttaacccat ctctgctggt ccgatccgta ctctggcgc ctccggatc  
601 aaagacttcc gcaaaatgtt ggctcattgc gaagccgtt ccccgattcg ccgttaccgtt  
661 actattgaag atgtggtaa ctctgcggca ttccctgtgtt ccgatcttc tgccggatc  
721 tccggtaag tggtccacgt tgacggcggt ttccagcattt ctgcaatgaa cgaactcga  
781 ctgaaataa

**Amino acid sequence (SEQ ID NO:192)**

1 mgflsgkril vtgvasklsi aygiaqamhr egaelaftyq ndklkgrvee faaqlgsdiv  
61 lqcdvaedas idtmfaelgk vwpkfdgfvh sigfapqdql dgdvnavtr egfkiahdis  
121 sysfvamaka crsmlnpgsa lltlsylgae raipnynvmg lakasleanv rymanamgpe  
181 gvrvnaisag pirtlaasi kdfrkmlahc eavtpirrtv tiedvgnssaa flcsdlsagi  
241 sgevvhvddgg fsiaamnele lk

**XdjA (NP 416279)****Nucleotide sequence (SEQ ID NO:193)**

1 atggatgcac tcgaaactatt gatcaatcgc cgttagcgcct cccgcttggc tgaaccggcg  
61 ccaacgggtt aacaactgca aaacatccctg cgtgcgggtt tgctgcgcgc ggaccataag  
121 tccatgcac cgtggcattt ttttgtattt gaagggggaaag ggcgcgagcg tttcagcgc  
181 gtactggAAC agggggcgat tgctgcgggtt agtgtatgaca aagctatcga caaaagccgt  
241 aatgcgcgcgt tccgcgcacc gctcatcata acgggtgggtt cggaaatgcga agagaatcat  
301 aaagtcccgc gctggaaaca gggaaatgtct gccggatgcg cggcatatggc gatgcaaatg  
361 gcagcagttt cccaggggtt tggcgccatc tggcgctgtt ggcattaaac tggaaatccg  
421 gtagtgcgtt aagcattcgg ttggcgctgtt caggataaaa tttgcgggttt tctcttacctc  
481 ggtacgcgcgc agctgaaagc atctacgtcg attaacgtcc cggacccgac ggcgtttgtt  
541 acttattttctt ga

**Amino acid sequence (SEQ ID NO:194)**

1 mdalellinr rsasrlaepa ptgeqlqnll ragmrapdhk smqpwhffvi egegrerfsa  
61 vleqgaiaag sddkaidkar napfraplri tvvakceenh kvprweqems agcavmamqm  
121 aavaqgffgi wrsgaltesp vvreafgcre qdkivgflyl gtpqlkasts invpdptpfv  
181 tyf)

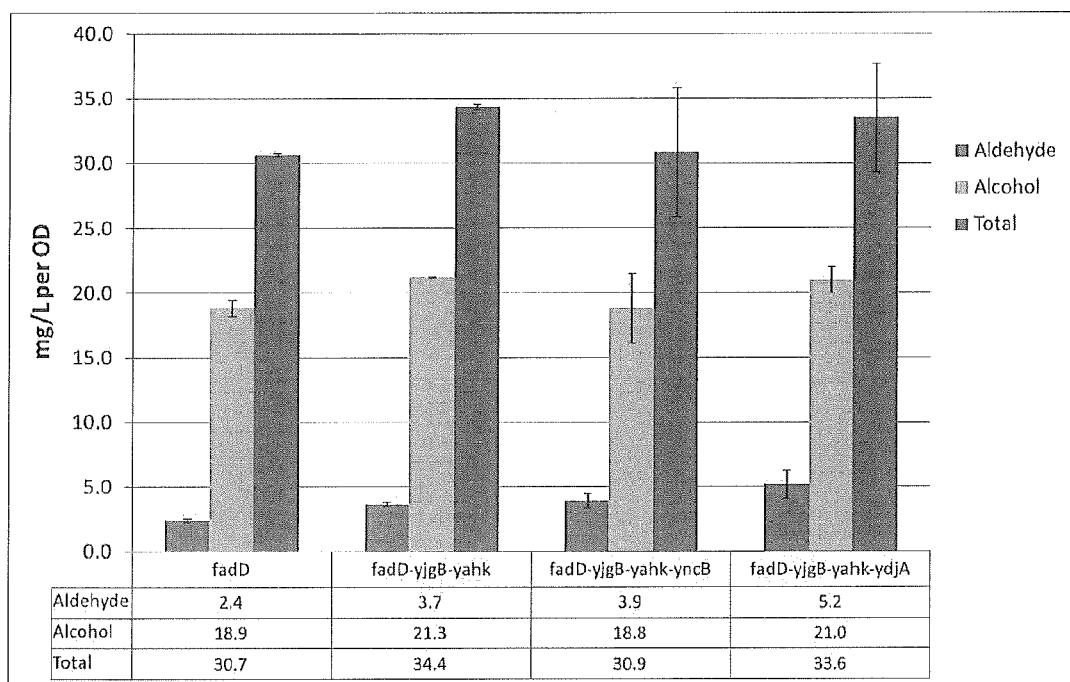


FIG. 11

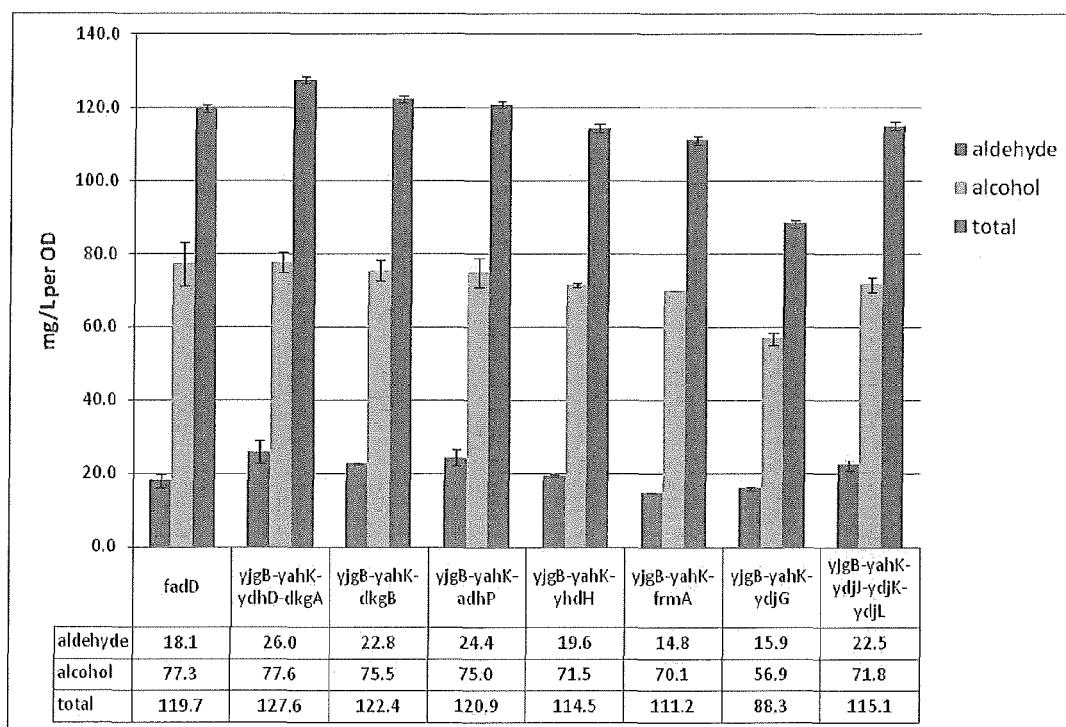


FIG. 12

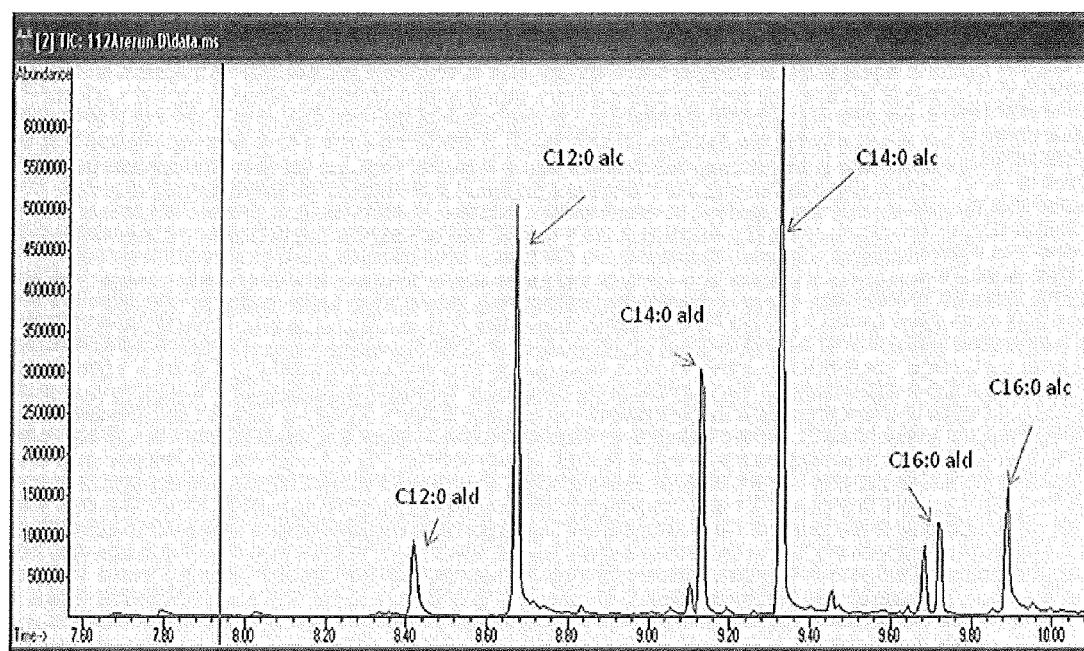


FIG. 13

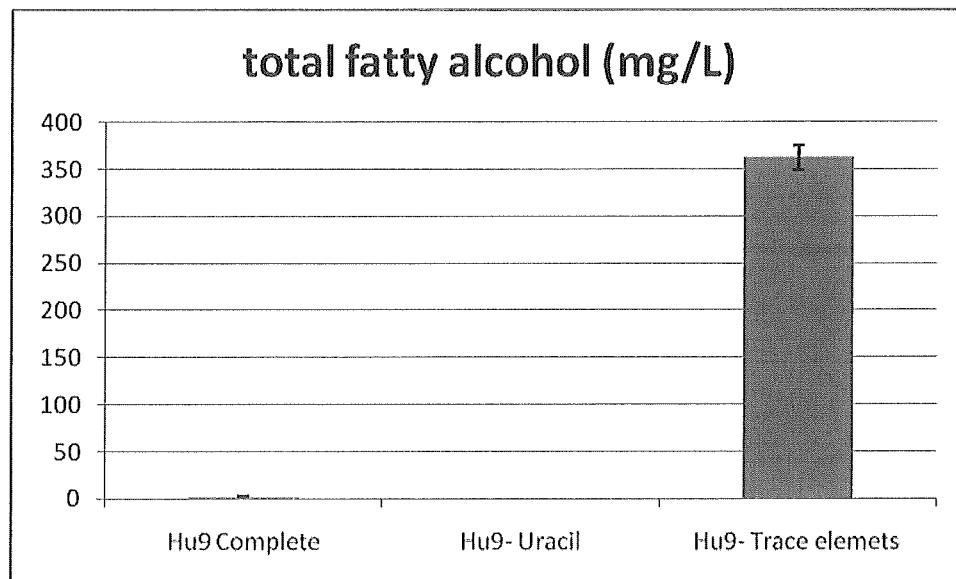
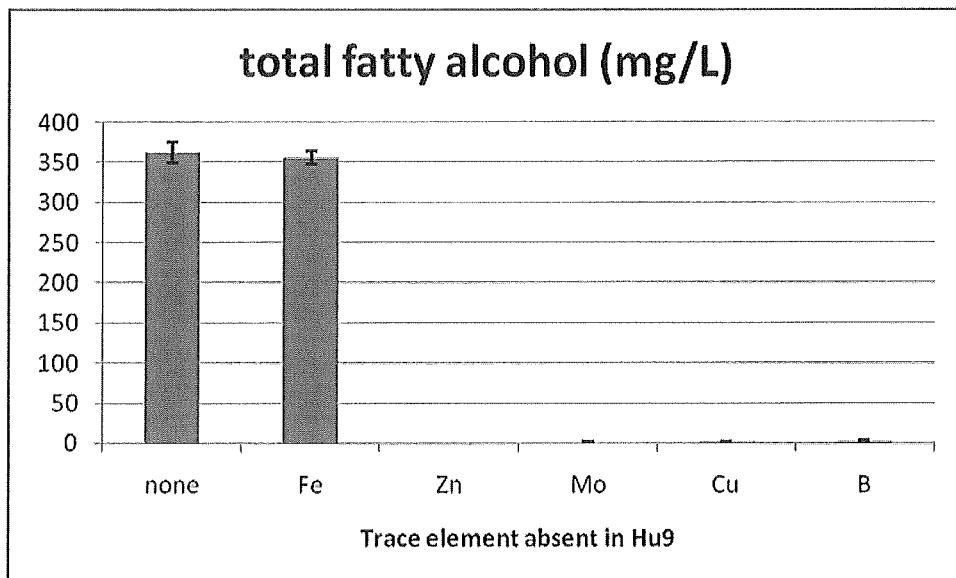
**A.****B.****FIG. 14**

FIG. 15A

Gene	Name	Nucleotide Sequence	Protein Sequence
<b>fabA</b>	beta-hydroxydecanoyl thioester dehydrase	atgGTAGATA AACGCGAAC CTATACAAAAA GAAGACCTTC TTGCCTCTGG TCGCGGTGAA CTGTTGGCG CTAAGGCC GCAATTGCCA GCACCGAAC TGCTGATGAT GGACCGTGTG GTCAAAATGA CCGAAACGGG TGGTAACCTC GACAAAGGAT ATGTTGAAGC AGAAACTGGAT ATCAATCAGG ATCTGTGGUT CTTCGGATGC CACTTTATTG GCGATCCGGT TATGCCGGGA TGCCTGGGCC TGGACGCAAT GTGGCACGCTG GTACGGTTCT ACCTCGGCTG GCTGGCGGC GAAGGTAAGA GCCCGCGCCT GGCGCTTGGC GAAGTGAAT TCACTGGTCA GGTACTGCCG ACACCGAAAAA AAGTGACCTA CCGTATTACAC TTTAACGCA TTGTTAACCG TCCTCTGATT ATGGGCCCTGG CGGATGCCGA AGTGTGCTT GATGGTCGTC TGATCTATAC CGCCAGCGAC CTGAAAGTCG GTCTGTTCCA GGATACGCTC GCCTCTGA (SEQ ID NO:243)	MVDKRESYTK EDLLASGRGE LFGAKGPQLP APNMLMMDRV VKMFETGGNF DKGYVVEAELD INPDWLWFFGC HFIGDPVMPG CLGLDAMWQL VCFYLIGWLGG EGKGGRALGVG EVKFTGQVLP TAKKVTYRIH FKRIVNRRLI MGLADGEVLV DGRLIYTASD LKVGGLFQDTS AF (SEQ ID NO:244)

**FIG. 15B**

<b>fabZ</b>	(3R)-hydroxymyristol acyl carrier protein dehydratase	ttgACTACTA ACACTCATACTCTGCAGATT GAAGAGATTTAGAACTCT TCCCGTTCT TACTGGTGGATTCCGTGCTG GATTTGAAGAAGTCGTT TCTGCGCGCACTAAAAATG TCTCTGTCAA TGAGCCATTC TTCCAGGGCCATTCCCTGG AAAACCGATTTCGGGGTG TGCTGATTCGGAAGCAATG GCACAGGCAA CAGGTATTCT GGCGTTAAAACCGTAGGAA AACTGGAACCGGTGAGCTG TACTACTTCGCTGGTATTGA CGAAGCGCGCTTCAAGCGCC CGCTCGTGCTGGCGATCAA ATGATCATGGAAAGTCACCTT CGAAAAAACCGCGCGCGCC TGACCCGTTAAGGGGTT GCTCTGGTCGATGGTAAAGT AGTTTGCAGAACCGATGA TGTGTGCTCGTAGCCGGAG GCCTGA (SEQ ID NO:245)	MTTNTHTLQI EEILELLPHR FPFLLVDRVL DFEEGRFLRA VKNVSVNEPF FQGHFPKGPI FPGVILIEAM AQATGILAFK SVGKLEPGEI YYFAGIDEAR FKRPVVPGDQ MIMEVTFEKT RRGLTRFKGV ALVDCKVVCE ATMMCARSRE A (SEQ ID NO:246)
-------------	---	---	--

FIG. 15C

<b>cysM</b>	cysteine synthase B (O-acetylserine sulphydrolase B)	gtgAGTACAT TAGAACAAAC AATAGGCAT ACGCCTCTGG TGAAGTTGCA GCGAATGGGG CCGGATAACG GCAGTGAAGT GTGGTTAAA CTGGAAGGCA ATAACC CGGC AGGTCGGTG AAAGATCGTG CGGCACITTC GATGATCGTC GAGGCCGGAAA AGCGCGGGAA ATTAAACCG GGTGAATGCT TAATCGAAC CACCA GTGGT AACACCGGCA TTGCGCTGGC AATGATTGCC GCGCTGAAAG GCTATCGCAT GAAATTGCTG ATGCCCGACA ACATGAGCCA GGAACCGCGT CCGGCGATGC GTGCTTATGG TGC GGA ACTG ATTCTTGTCA CCAAAGAGCA GGGCATGGAA GGTGC CCGCG ATCTGGCGCT GGAGATCGCG AATCGTGGCG AAGGAAAGCT GCTCGATCAG TTCAATAATC CCGATAACCC TTATGCCAT TACACCACCA CTGGGCCGGA AATCTGGCAG CAAACCGGCG GGC GCA TCAC TCATTTGTC TCCAGCATGG GGACGACCGG CACTATCACC GGCGTCTCAC GCTTTATGCG CGAACATCC AAACCGGTGA CCATTGT CGG CCTGCAACCG GAAGAGGGCA GCAGCATTCC CGGCATT CGC CGCTGGCTA CGGAATATCT GCCGGGGATT TTCAACCGCTT CTCTGGT CGGA TGAGGTGCTG GATATT CATC AGCGCGATGC GGAAAACACC ATGCCCGAAC TGGCGGTGCG GGAAGGAATA TTCTGTGGCG TCAGCTCCGG CGGC CGCGGTT GCCGGAGCAC TGCGGGTGGC AAAAGCTAAC CCTGACCGCG TGGTGGTGGC GATCATCTGC GATCGTGGCG ATCGCTACCT TTCTACCGGG GTGTTGGGG AAGACGATT TAGCCAGGGG GCGGGGATTT AA (SEQ ID NO: 247)	MSTLEQTIGN TPLVKLORMG PDNGSEVWLK LEGNNPAGSV KDRAALSMIV EAEKRG EIKP GDVLI EATSG NTGIALAMIA ALKGYRMKLL MPDNMSQERR AAMRAYGAEL ILVTKEQGME GARDIALEMA NRGECKLLDQ FNNPDNPYAH YTTTGP EIWO QTGGRITHFV SSMGTGTT GVSRMREQS KPVTIVGLQP EEGSSIPGIR RWPT EYLPGI FNASLVDEVL DIHQ RDAENT MRELAVREGI FCGVSSGGAV AGALRVAKAN PD AVVVAIIC DRGD RYLSTG VFGE EHFSQG AGI (SEQ ID NO: 248)
-------------	--	---	---

FIG. 15D

<b>maoC</b>	fused aldehyde dehydrogenase/enoyl-CoA hydratase	atgCAGCAGT TAGCCAGTT CTTATCCGGT ACCTGGCAGT CTGGCCGGG CCCTAGCCGT TTGATTCAAC ACGCTATTAG CGGCGAGGCG TTATGGGAAG TGACCAGTGA AGGTCTTGAT ATGGCGGCTG CCCGCCAGTT TGCCATTGAA AAAGGTGCC CCGCCCTTCG CGCTATGACC TTTATCGAAC GTGCGGCAT GCTTAAGCG GTCGCTAAC ATCTGCTGAG TGAAAAAGAG CGTTTCTATG CTCTTCTGC CCAAACAGGC GCAACCGGG CAGACAGTIG GGTGATATT GAAGGTGGCA TTGGGACGTT ATTACTTAC GCCAGCCTCG GTAGCCGGGA CCTCCCTGAC GATA CGCTGT GGCGGAAGA TGAATTGATC CCCTTATCGA AAGAAGGTGG ATTTGCCCG CGCCATTAC TGACCTAAA GTCAGGGCTG GCAGTGCATA TTAACGCCCT TAACTTCCCC TGCTGGGAA TGCTGGAAAA GCTGGCACCA ACCTGGCTGG GCGGAATGCC AGCCATCATC AAACCACCTA CCGCACCGC CCAAC TGA CAGGGATGG TGAAATCAAT TGTGATAGT GGTCTTGTTC CCGAAGGC AATTAGCTG ATCTGGGTA GTGCTGGCGA CTTGTTGGAT CATCTGGACA GCCAGGATGT GGTGA CTTAC CGGGGCTAG CGGGCACCCG ACAGATGCTG CGAGTTCAAGC CAAATATCGT CGCCAAATCT ATCCCTTCA CTATGGAAAGC TGATCCCTG AACTGCTGCG TACTGGGCGA AGATGTCACC CCGGATCAAC CGGACTTGC GCTGTTTATT CGTGAAGTTG TGCGTGAGAT GACCACAAAA GCCGGGCAA AATGTCAGGC AATCCGGCG ATTATTGTCG CCGAGGCATT GGTTAATGCT GTCAGTGATG CTCTGGTTGC GCGATTACAG AAAGTCGTGG TCGGTGATCC TGCTCAAGGAA GGCCTGAAAA TGGGCGCACT GGTAAATGCT GAGCAGCGTG CCGATGTGCA GGAAAAGTG AACATATTGC TGGCTGCAGG ATGCGAGATT	MQQLASFLSG TWOSGRGRSR LIHHAIISGEA LWEVTSEGLD MAAARQFAIE KGAPALRAMT FIERAAMLKA VAKHLLSEKE RFYAIJSAQTG ATRADSWVDI EGGIGTLFTY ASLGSRELPD DTLWPEDELT PLSKEGFFAA RHLLTSKSGV AVIHINAENFP CWGMLEKLAP TWLGGMPAII KPATATAQLT QAMVKSIIVDS GLVPEGAISL ICGSADLD HLDSDQDVTF TGSATGQML RVQPNIVAKS IPFTMEADSL NCCVLGEDVT PDQPEFALFI REVVEREMTTK AGQKCTAIRR IIVPQALVNA VSDAIVARLQ KVVVGDPQA GVKM GALVNA EQRADVQEKV NTLLAAGCEI RLGGQADLSA AGAFFPPTLL YCPQPDETPA VHATEAFGPV ATLMPAQNQR HALQLACAGG GSLAGTLVTA DPQIARQFIA DAARTHGRIO ILNEESAKES TGHGSPLPQL VHGGPGRAGG GEELGGLR KHYMORTAVQ GSPTMLAAIS KQWVRGAKVE EDRIHPFRKY FEELQPGDSL
-------------	--	---	--

FIG. 15E

		CGCCTCGGTG GTCAGGGCGGA TTTATCTGCT GCGGGTGCCT TCTTCCCGCC AACCTTATTG TACTGTCCGC AGCCCGATGA AACACCGGCG GTACATGCAA CAGAACGCTT TGGCCCTGTC GCAACGCTGA TGCCAGCACA AAACCAGCGA CATGCTCTGC AACTGGCTTG TGCAGGGCGGC GGTAGCCTTG CGGGAACGCT GGTGACGGCT GATCCGAAA TTGCGCGTCA GTTTATTGCC GACGCGGCAC GTACGCATGG GCGAATTCAAG ATCCTCAATG AAGAGTCGGC AAAAGAATCC ACCGGCCATG CCTCCCCACT GCCACAACGT GTACATGGTG GGCCTGGTCG CGCAGGGAGGC GGTGAAGAAT TAGGCGGTTT ACGAGGGGTG AAACATTACA TGCAGCCAAC CGCTGTTTCAG GGTAGTCGCA CGATGCTTGC CGCTATCAGT AAACAGTGGG TGCAGCGTGC GAAAGTCGAA GAAGATCGTA TTCATCCGTT CCGCAAATAT TTTGAGGAGC TACAACCAGG CGACAGCCTG TTGACTCCCC GCCGCACAAT GACAGAGGCC GATATTGTTA ACTTTGCTTG CCTCAGCGGC GATCATTCT ATGCACATAT GGATAAGATT GCTGCTGCCG AATCTATTT CGGTGAGCGG GTCTTGCTATG GGTATTTTGT GCTTTCTGCG GCTGCGGGTC TGTTTGTGCA TGCCGGTGTG GGTCCGGTCA TTGCTAACTA CGGGCTGGAA AGCTTGCCTT TTATCGAACCC CGTAAAGCCA GGCGATACCA TCCAGGTGCG TCTCACCTGT AAGCGCAAGA CGCTGAAAAAA ACACCGTAGC GCAGAAGAAA AACCAACAGG TGTGGTGCAGA TCGGGCTGTAG AGGTATTCAA TCAGCATCAA ACCCCGGTGG CGCTGTATTG AATTCTGACG CTGGTGGCCA GGCAGCACGG TGATTTGTC GATTA (SEQ ID NO:249)	LTPRRTMTEA DIVNFACLSG DHFYAHMDKI AAAESIFGER VVHGYFVLSA AAGLFVDAGV GPVIANYGLE SLRFIEPVKP GDTIQVRLTC KRKTLKKQRS AEEKPTGVVE WAVEVFNQHQ TPVALYSILT LVARQHGDFV D (SEQ ID NO:250)
--	--	---	---

FIG. 15F

Source	Genbank Accession Number
<i>Shigella</i> sp. D9	ZP_05432652
<i>Citrobacter youngae</i> ATCC 29220	ZP_04561391.1
<i>Salmonella enterica</i>	YP_001570967.1
<i>Escherichia fergusonii</i> ATCC 35469	YP_002382254.1
<i>Klebsiella pneumoniae</i> NTUH-K2044	YP_002918743.1
<i>Enterobacter cancerogenus</i> ATCC 35316	ZP_03281954.1
<i>Cronobacter turicensis</i>	CBA29728.1
<i>Erwinia pyrifoliae</i> Ep1/96	YP_002649242.1
<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i> PC1	YP_003018119.1
<i>Dickeya dadantii</i> Ech703	YP_002987184.1
<i>Edwardsiella ictaluri</i> 93-146	YP_002932813.1
<i>Providencia alcalifaciens</i> DSM 30120	ZP_03317956.1
<i>Yersinia kristensenii</i> ATCC 33638	ZP_04624337.1
<i>Photorhabdus asymbiotica</i>	YP_003041580.1
<i>Pantoea</i> sp. At-9b	ZP_05728924.1
<i>Actinobacillus succinogenes</i> 130Z	YP_001344737.1

FIG. 15G

<i>Mannheimia succiniciproducens</i> MBEL55E	YP_088386.1
<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	NP_245421.1
<i>Haemophilus somnus</i> 129PT	YP_719117.1
<i>Proteus mirabilis</i> HI4320	YP_002150544.1
<i>Sodalis glossinidius</i> str. 'morsitans'	YP_454706.1
<i>Candidatus Blochmannia pennsylvanicus</i> str. BPEN	YP_277927.1
<i>Aggregatibacter aphrophilus</i> NJ8700	YP_003007342.1
<i>Vibrio cholerae</i> MZO-3	ZP_01958381.1
<i>Baumannia cicadellinicola</i> str. Hc ( <i>Homalodisca coagulata</i> )	YP_588853.1
<i>Vibrionales bacterium</i> SWAT-3	ZP_01815187.1
<i>Aliivibrio salmonicida</i> LFI1238	YP_002262988.1
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	YP_001141819.1
<i>Wigglesworthia glossinidiae</i> endosymbiont of <i>Glossina brevipalpis</i>	NP_871303.1
<i>Glaciecola</i> sp. HTCC2999	ZP_03560821.1
<i>Alteromonas macleodii</i> ATCC 27126	ZP_04714556.1

FIG. 16A

Gene	Name	Nucleotide Sequence	Protein Sequence
<b>fabB</b>	B-ketoacyl synthase/ 3-oxoacyl-[acyl- carrier-protein] synthase I	atgAAACGTG CAGTGATTAC TGGCCTGGGC ATTGTTTCCA GCATCGGTAA TAACCAGCAG GAAGTCCTGG CATCTCTGCG TGAAGGACGT TCAGGGATCA CTTTCTCTCA GGAGCTGAAG GATTCCGGCA TGCCTAGCCA CGTCTGGGGC AACGTAACAC TGGATACAC TGGCTCATT CACCCCAAAG TTCTGCCCTT TATGAGCGAC GCATCCATT ATGCATTCTT TTCTATGGAG CAGGCAATCG CTGATCCGGG CCTCTCTCCG GAAGCTTACCC AGAATAACCC CGCGCTTGGC CTGATTGCCAG GTTCCGGCG CGGCTCCCCG CGTTTCCAGG TGTTCGGCGC TGACCCAATG CGCGGCCCGC CGGGCCTGAA ACCGGTTGGC CCGTATGTGG TCACCAAAGC GATGGCATCC GGCGTTCTG CCTGCCCTCG CACCCCGTTT AAAATCATG GCGTTAACTA CTCCATCAGC TCCCGTGTG CGACTTCCGC ACACTGTATC GGTACCGAG TACAGCAGAT CCAACTGGGC AAACAGGACA TCGTGTGTTGC TGGCGGCCGC GAAGAGCTGT GCTGGAAAT GGCTTGCAGA TTCGACGCAA TGGTTCGCT GTCTACTAAA TACAACGACA CCCCGGAAAA AGCCTCCCCT ACTTACGACG CTCACCGTGA CGGTTTCGTT ATCGCTGGCG CGGGCGGTAT GCTAGTGGTT GAAGAGCTGG AACACCGCCT GGCGCGTGGT GCTCACATCT ATGCTGAAAT CGTTGGCTAC GGCGCAACCT CTGATGGTGC AGACATCGTT GCTCCGTCTG GCGAAGGCCGC AGTACGCTGC ATGAAGATGG CGATCGATGG CGTTGATACC CCAATCGATT ACCTGAACCTC CCACCGTACT TCGACTCCGG TTGGCGACGT GAAAGAGCTG GCAGCTATCC GTGAAGTGTG CGGCAGATAAG AGCCCAGGCGA TTCTGCAAC CAAAGCCATG ACCGGTCACT CTCTGGCGC TGCTGGCGTA CAGGAAGCTA TCTACTCTCT	MKRAVITGLG IVSSIGNNQQ EVLASLREGR SGITFSQELK DSGMRSHVWG NVKLDTTGLI DRKVVRFMSD ASIYAFLSME QAIADAGLSP EAYQNPNRVG LIAGSGGGSP RFQVFGADAM RGPRGLKAVG PYVVTKAMAS GVSACLATPF KIHGVNYSIS SACATSACI GNAVEQIQLG KQDIVFAGGG EELCWEMACE FDAMGALSTK YNDTPEKASR TYDAHRDGTV IAGGGGMVVV EELEHALARG AHIYAEIVGY GATSDGADMV APSGEHAVRC MKMAMHGVDT PIDYLNSHGT STPVGDVKEL AAIREVFGDK SPAISATKAM TGHSLGAAGV QEAIYSLLML EHGFIAPSIN IEELDEQAAG LNIVTETTDR ELTTVMSNSF GFGGTNATLV MRKLKD (SEQ ID NO:252)

FIG. 16B

		GCTGATGCTG GAACACGGCT TTATCGCCCC GAGCATCAAC ATTGAAGAGC TGGACGAGCA GGCTGCGGGT CTGAACATCG TGACCGAAAC GACCGATCGC GAACTGACCA CCGTTATGTC TAACAGCTTC GGCTTCGGCG GCACCAAACGC CACCGTGGTA ATGCGCAAGC TGAAACATTA A (SEQ ID NO:251)	
<b>fabF</b>	3-oxoacyl-[acyl-carrier-protein] synthase II	gt.gTCTAAGC GTCGTGTAGT TGTGACCGA CTGGCATGT TGTCTCCGT CGGAAATACC GTAGAGTCTA CCTGGAAAGC TCTGCTGCC GGTCAAGAGT GCATCAGCCT AATCGACCAT TTCGATACTA GCGCTATGC AACGAAATTG CCGGCTTAG TAAAGGATTT TAACTGTGAG GACATTATCT CGCGAAAGA ACAGCGCAAG ATGGATGCCT TCATTCAATA TCCAATTGTC GCTGGCGTTG AGGCCATGCA GGATTCTGGC CTTGAAATAA CGGAAGAGAA CGCAACCCGC ATTGGTGCCG CAATTGGCTC CGGGATTGGC CGCCTCGGAC TGATCGAAGA AAACACACA TCTCTGATGA ACGGTGGTCC ACGTAAGATC AGCCCATTCT TCGTTCCCTC AACGATTGTG AACATGGTGG CAGGTCACT GACTATCATG TATGCCCTGC GTGGCCCGAG CATCTCTATC GCGACTGCC GTACTTCCGG CGTGCACAAAC ATTGGCCATG CTGCGCGTAT TATCGCGTAT GGCGATGCTG ACGTGATGGT TGCAGGTGGC GCAGAGAAAG CCAGTACGCC GCTGGCCGTT GGTGGTTTG GCGCGGCACG TGCATTATCT ACCCGCAATG ATAACCCGCA AGCGGGCAGC CGCCCGTGGG ATAAGAGAGC TGATGGTTTC GTACTGGCG ATGGTGGCGG TATGCTGGTA CTTGAAGAGT ACGAACACAGC GAAAAAACGC GGTGCGAAAA TTTACCCCTGA ACTCCTCGCC TTTGGTATGA GCAGCGATGC TTATCATATG ACGTCACCGC CAGAAAATGG CGCAGGCGCA	MSKRRVVVTG LGMLSPVGNT VESTWKALLA GQSGISLIDH FDTTSAYATKF AGLVKDFNCE DIISRKEQRK MDAFIQYQGIV AGVQAMQDSG LEITEENATR IGAAIGSGIG GLGLIEENHT SLMNGGPRKI SPFFVPSTIV NMVAGHLTIM YGLRGPSISI ATACTSGVHN IGHAARIJAY GDADVMVAGG AEKASTPLCV GGFGAARALS TRNDNPQAAS RPWDKERDGF VLGDGAGMLV LEEYEHAKKR GAKIYIAELVG FGMSSDAYHM TSPPENGAGA ALAMANALRD AGIEASQIGY VNAHGTSTPA GDKAEAQAVK TIFGEAASRV LVSSTKSMTG HLLGAAGAVE SIYSILALRD QAVPPTINLD NPDEGCBLDF VPHEARQVSG MEYTLICNSFG FGGTNGSLIF KKI (SEQ ID NO:254)

FIG. 16C

		GCTCTGGCGA TGGCAAATGC TCTGGGTGAT CCAGGCATTG AAGCGAGTCA GATTGGCTAC GTTAACCGCGC ACGGTACTTC TACGCCGCT GGGGATAAAG CTGAAGCCGAA GGCGGTGAAA ACCATCTCG GTGAAGCTGC AAGCCGTGTG TTGGTAAGCT CCACGAAATC TATGACCGGT CACCTGTAG GTGCGGCGGG TGCAGTAGAA TCTATCTACT CCATCCITGGC GCTGCGCGAT CAGGCTGTTG CGCCAACCAT CAACCTGGAT AACCCGGATG AAGGTTGCGA TCTGGATTTC GTACCGCACG AAGCCCCGCA GGTTAGCGGA ATGGAATACA CTCTGTGTA CTCCTTCGGC TTCGGTGGCA CTAATGGTTC TTTGATCTT AAAAAGATCT AA (SEQ ID NO:253)	
<b>fadJ</b>	fused enoyl-CoA hydratase and epimerase and isomerase/3-hydroxyacyl-CoA dehydrogenase	atqGAAATGA CATCAGCGTT TACCCCTTAAT GTTCTGTCTGG ACAACATGTC CGTTATCACC ATCGACGTAC CGGGTGGAGAA AATGAATACC CTGAAGGCAG AGTTTGCTC GCAGGGTGC GCCATTATTA AGCAACTCCG TGAAAACAAA GAGTTGCGAG GCGTGGTGTG TGTCTCCGCT AAACCGGACA ACTTCATTGC TGGCGCAGAC ATCAACATGA TCGGCAACTG CAAAACGGCG CAAGAAGCGG AAGCTCTGGC GCGGCAGGGC CAACAGTTGA TGGCGGAGAT TCATGCTTTC CCCATTCAAGG TTATCGCGC TATTCACTGGC GCTTGCTTGG CTGGTGGGCT GGAGTTGGCG CTGGCGTGCC ACAGTCGCGT TTGTACTGAC GATCTAAAA CGGTGCTCGG TTTGCCCTGAA CTACAACCTG GATTGTTTAC CGGTTCAAGGC GGCAACCCAGC GTTTACCGCG TCTGATAGGC GTCAGCACAG CATTAGAGAT GATCCTCACCC GGAAAACAAC TTCGGGGCGAA ACAGGCATTA AAGCTGGGGC TGTTGGATGA CGTTGTTCCG CACTCCATTC TGCTGGAAGC CGCTGTTGAG CTGGCAAAGA AGGAGCGCCC ATCTTCCCGC CCTCTACCTG	MEMTSAFTLN VRLDNTIAVIT IDVPGEKMNT LKAEFASQVR AIIKQLRENK ELRGVVVFSA KPDNFIAAGAD INMIGNCKTA QEAEALARQG QQLMAEIHAL PIQVIAIHG ACLGGLEJA LACHGRVCTD DPKTVLGLPE VQLGIJLPGS GTQRPLRILIG VSTALEMILT GKQLRAKQAL KLGLVDDVVP HSILLEAAVE LAKKERPSSR PLPVVERILA GPLGRALLFK MVGKKTEHK QGNYPATERI LEVVEETGLAQ GTSSGYDAEA RAFGELAMTP QSQALRSIFF ASTDVKKDPG SDAPPAPLNS VGILGGGLMG

FIG. 16D

		TACGCGAGCG TATTCTGGCG GGGCCGTAG GTCGTGCGCT GCTCTTCAAA ATGGTCGGCA AGAAAACAGA ACACAAAACT CAAGGCATT ATCCGGCGAC AGAACCCATC CTGGAGGTTG TTGAACCGG ATTAGCGCAG GGCACAGCA GCGGTTATGA CGCCGAAGCT CGGGCGTTG GCGAACTGGC GATGACGCCA CAATCGCAGG CGCTGGTAG TATCTTTT GCCAGTACGG ACGTGAAGAA AGATCCCAGC AGTGATGCGC CGCCTGCGCC ATTAAACAGC GTGGGGATT TAGGTGGTGG CTTGATGGC GGCGTATTG CTTATGTCAC TGCTTGTAAA GCGGGGATT CGGTCAAGAT TAAAGATATC AACCCGCAGG GCATAAAATCA TGGCCTGAAG TACAGTTGGG ATCAGCTGGA GGGCAAAGTT CGCCGTGTC ATCTCAAAGC CAGCGAACGT GACAAACAGC TGGCATTAAT CTCCCGAACG ACGGACTATC GCGGCATTGC CCATCGCGAT CTGATTATTG AAGCGGTGTT TGAAAATCTC GAATTGAAAC AACAGATGGT GGCGGAAGTT GAGCAAATT GCGCGCTCA TACCATCTTT GCTTCGAATA CGTCATCTTT ACCGATTGGT GATATCGCCG CTCACGCCAC GCGACCTGAG CAAGTTATCG GCCTGCATT CTTCAGTCCG GTGGAAAAAA TCCCGCTGGT GGAGATTATT CCTCATGCGG GGACATCGGC GAAACCATC GCTACCACAG TAAAACGGC GAAAAAACAG GGTAAAACGC CAATTGTCGT GCGTGACAAA GCCGTTTTT ACGTCAATCG CATCTTAGCG CCTTACATTA ATGAAGCTAT CCGCATGTTG ACCCAAGGTG AACGGGTAGA GCACATTGAT GCCGCGCTAG TGAAATTGG TTTCCCGGTA GGCCCAATCC AACTTTGGA TGAGGTAGGA ATCGACACCG GGACTAAAT TATTCCCTCTA CTGGAAGCCG CTTATGGAGA ACGTTTAGC GCGCCTGCAA ATGTTGTTTC	GGIAYVTACK AGIPVRIKDI NPQGINHALK YSWDQLEGKV RRRHILKASER DKQLALISGT TDYRGFAHRD LIIEAVFENL ELKQQMVAEV EQNCAAHТИ ASNITSSLPIG DIAAHATRPE QVIGLHFFSP VEKMPLVIEII PHAGTSAQTI ATTVKLAKKQ GKTPIVVRDK AGFYVNRLA PYINEAIRML TQGERVEHID AALVKFGFPV GPIQLLDEVG IDTGTKIIPV LEAAYGERFS APANVSSIL NDDRKGKRKNG RGFYLYGQKG RKSKKQVDPA IYPLIGTQGQ GRISAPQVAE RCVMILMLNEA VRCVDEQVIR SVRDGDIGAV FGIGFPPFLG GPFYRIDSLG AGEVVAIMQR LATQYGSRFT PCERLVEMG RGESFWKT TA TDLO (SEQ ID NO: 256)
--	--	--	---

FIG. 16E

		TTCAATTG AACGACGATC GCAAAGGCAG AAAAAATGGC CGGGGTTCT ATCTTATGG TCAGAAAGGG CGTAAAAGCA AAAAACAGGT CGATCCGCC ATTTACCCGC TGATGGCAC ACAAGGGCAG GGGCGAATCT CCGCACCGCA GGTTGCTGAA CGGTGTGTGA TGTTGATGCT GAATGAAGCA GTACGTTGTG TTGATGAGCA CGTTATCCGT AGCGTGCGTG ACGGGGATAT TGGCGCGTA TTTGGCATTG GTTTCCGCC ATTTCTCGGT GGACCGTTCC GCTATATCGA TTCTCTCGCC GCAGGGCAAG TGGTTGCAAT AATGCAACGA CTTGGCACCGC AGTATGGTTC CCGTTTTAAC CCTTGGCAGC GTTTGGTCGA GATGGGCGCG CGTGGGGAAA GTTTTGGAA AACAACTGCA ACTGACCTGC AATAA (SEQ ID NO:255)	
xerC	site-specific tyrosine recombinase	atgACCGATT TACACACCAGA TGTACAACGC TACCTACGTT ATCTGAGCGT GGAGCGCCAG CTTAGCCCGA TAACCCCTGCT TAACTACCAAG CCTCAGCTTG AGGCAGATCAT CAATTTCGCC AGCGAAAACG GCCTGCAAAG CTGGCAGCAA TGTGATGTGA CGATGGTGCG CAATTTCGCT GTACGCAGTC GCCGTAAGG GCTGGGAGCA GCAAGTCTGG CGTTACGGCT TTCTCGCCTA CGTAGCTTT TTGACTGGCT GGTCAGCCAG AACGAACCTCA AAGCTAACCC GGCAGAAAGGT GTTTCGGCAC CGAAAGCGCC GCGTCATCTG CCGAAAAACA TCGACGTGCA CGATATGAAT CGGCTGCTGC ATATTGATAT CAATGATCCC CTCGCTGTAC GCGACCGTGC AATGCTGGAA GTGATGTACG GCGCGGGTCT GCGTCCTTCT GAGCTGGTGG GGCTGGATAT TAAACACCTC GACCTGGAGT CTGGTGAAGT GTGGGTTATG GGGAAAGGCA GCAAAGAGCG CCGCCCTGCCG ATTGGTGCAG ACGCTGTGGC GTGGATTGAG CACTGGCTTG ATTTGCGCGA CCTGTTGGT	MTDLHTDVER YLRYLSVERQ LSPITLJNYQ RQLEAIINFA SENGLQSWQQ CDVTMVRNFA VRSRRKGLGA ASLALRLSAL RSFFDWLVSQ NELKANPAKG VSAPKAPRHL PKNIDVDDMN RLLDIDINDP LAVRDRAMLE VMYGAGLRLS ELVGLDIKHL DLESGEVWVM GKGSKERRLP IGRNAVAWIE HWLDLRLDFG SEDDALFLSK LGKRISARNV OKRFAEWGIK QGLNNHHVPH KLRHSFATHM LESSGDLRGV OELLGHANLS TTQIYTHLDF QHLASVYDAA HPRAKRGK (SEQ ID NO:258)

FIG. 16F

		AGCGAAGACG ACGCGCTTT TCTGTCGAAA CTGGGCAAGC GTATCTCCGC GCGTAATGTG CAGAAACGCT TTGCCGAATG GGGCATAAAA CAAGGGCTGA ATAATCACGT TCATCCGCAT AAATTACGTC ACTCGITTCGC CACGCATATG CTGGAGTCGA GCGGCGATCT TCGTGGTGTG CAGGAGCTGC TGGGTCAATGC CAACCTCTCC ACCACGCAA TCTATACTCA TCTTGATTTT CAACACCTTG CCTCGGTGTA CGATGCGGCG CATCCACGCG CCAAACGGGG GAAATAA (SEQ ID NO:257)	
yqeF	predicted acyltransferase	atgAAAGACG TTGTGATTGT CGGGGCGTTA CGGACACACTA TCGGCTGCTT TCGTGGTGC TTAGCCGGTC ATTCCGGCGT GGAACCTGGT AGTCTGGTC TGAAAGCGTT AATAGAACGT ACC GGCGTTC CTGCATATGC GGTGGATGAA GTAATTCTTG GTCACCTGTT GACTGCAGGG GCAGGGCAGA ATCCGGCAAG GCAATCGGCT ATTAAGGTG GTCCTGCCTAA TAGCGTTTCT GCAATCACTA TTAATGACGT TTGCGGTCC GGGCTTAAAG CACTGCATCT GGCTACTCAG GCGATAACAGT GTGGCGAGGC TGATATTGTC ATCGCCGGTG GCCAGGAAAAA CATGAGCCGC GCACCCACATG TTCTGACTGA TAGCCGCACC GGTGCACAGC TTGGCAATAG CCAGTTGGTT GACAGTCITG TGCATGATGG GTTGTGGGAT GCCTTCAATG ATTATCATAT TGGTGTCA GCCGAAAATC TGGCTCGCGA ATATGGCATC AGCCGTCAGT TGCAGGATGC TTACGCACCT AGCTCGAAC AAAAACGCGC AGCGGGCATT GACGCCGGAC GATTTAAAGA TGAGATCGTC CCGGTAATGA CCCAAAGTAA CGGGCAGACG TTGGTTGTTG ATACCGATGA ACAGCCACGC ACTGACGCCA GCGCAGAAGG CTTAGCCCGT TAAATCCCT CATTTGATAG TCTCGGTCT GTGACACCGG GTAATGCATC	MKDVVIVGAL RTPIGCFRGA LAGHSAVELG SLVVKALIER TGVPAYAVDE VILGQVLTAG AGONPARQSA IKGLPNSVS AITINDVCGS GLKALHLATQ AIQCCEADIV IAGGQENMSR APHVLTDSRT GAQLGNSQLV DSLVDGLWD AFNDYHIGVT AENLAREYGI SRQLQDAYAL SSQQKARAII DAGREKDEIV PVMTQSNQQT LVVDTDEQPR TDASAEGLAR LNPSFDLGS VIAGNASSIN DGAAAVMMMS EAKARALNLP VIARIKFAS VGVDPALMGI APVYATRRCL ERVGWQLAEV DLIEANEAF AQALSVGKML EWDERRVNVN GGAIALGHPI GASGCRILVS LVHEMVKRNA

FIG. 16G

		ATCCATAAAC GATGGCGCAG CTGCAGTAAT GATGATGAGC GAAGCCAAAG CACGACCGTT GAATTTACCC GTGCTGGCCC GCATTCCGC ATTTGCCAGC GTTGGTGTAG ATCCGGCATT GATGGGAATT CGCCGGTGT ATCCGACCCG CCGTTGCCCTG GAGCGTGTAG GCTGGCAGTT GGCTGAAGTC GATCTTATCG AGGCTAATGA AGCGTTTGCT GCACAGGCAGC TTTCGGTTGG CAAGATGCTT GAGTGGCATG AGCGTCGGGT CAATGTCAT GGTGGCCGCA TCGCACTCGG TCACCCGATA GGCGCTTCCG GTTGCCAAT CCTGGTTTCT CTGGTTCATG AAATGGTGAA ACGTAATGCC CGCAAAGGAC TGGCAACGCT TTGTATCGGC GGGGGCCAGG GTGTGGCATT GACCATTGAA CGTGACGAAT AG (SEQ ID NO:259)	RKGLATLCIG GGQGVALTIE RDE (SEQ ID NO:260)
murQ	predicted PTS component	atgCAATTG AAAAGATGAT TACTGAAGGC TCGAACACCG CCTCGGTGA AATTGACCGC GTATCGACCC TGGAAATGTG CCGGATTATC AACGATGAAG ATAAAAACCGT ACCGCTTGCC GTTGACCGCG TACTGCCGGA TATGCCGCG GCGATCGATG TTATCCACGC CCAGGTAAGC GGCGCGGGC GTCTGATTAA CCTCGGTGCG GGAACATCCG GTCGCTCGGG GATTCTGGAT GCCAGCGAAT GTCCGCCAC CTACGGCGTG AAACCGGGTC TGGTGGTGG TTTGATTGCT GGCGGCGAAT ATGCATTCA GCACGCGGTG GAAGGCCGG AAGATAGCCG GGAAGGGCGT GTTAATGATC TGAAAAATAT TAATTTAACG GCACAGGATG TGGTGGTGG CATTGCTGCC ACCGGTGCGA CGCCGTATGT GATTGCCCGA CTGGAAATACG CACGCCAGCT CGGCTGCCGC ACAGTGGGAA TTTCTGTAA TCCGGGGAGC GCCGTTTCAA CCACCGCTGA GTTGCCATT ACACCGATTG TAGGTGCCGA AGTTGTTACC GGTTCTTCGC	MDFERKMITEG SNTASAEIDR VSTLEMCRRI NDEDKTVPLA VERVLPDIAA AIDVIHAQVS GGGRILYLG GTSGRRLGILD ASECPPTYGV KPGLVVGLIA GGEYAIQHAV EGAEDSREGG VNDLKNINLT AQDVVVGIAA SGRTPYVIAG LEYARQLGCR TVGISCNPGS AVSTTAEFAI TPIVGAEVVT GSSRMKAGTA OKLVLNMLST GLMIKSCKVF GNLMVDVVAT NEKLHVRQVN IVKNATGCSA FQAEEAALIAC ERNCKTAIVM VLKNLDAAEA KKRLDQHGGF IRQVLDKE

FIG. 16H

	GGATGAAAGC AGGTACAGCG CAGAAACTGG TGCTCAATAT GCTTTCCACC GGGCTGATGA TTAAATCCCG CAAACTGTTTC GGCAACCTGA TGGTCGATGT GGTCCGCACC AACGAAAAAC TGCATGTGCG ACAGGTCAAT ATTGTTAAAA ACGCCACCGG ATGTAGCGCA GAGCAAGCGG AAGCAGCGTT AATTGCTTGC GAGCGCAACT GTAAAACGGC CATTGTGATG GTGCTGAAAA ATCTCGATGC CCCAGAAGCT AAAAAACGCC TGGATCAACA CGGCAGCTTT ATTGTCAGG TTTAGACAA GGAATAA (SEQ ID NO:261)	(SEQ ID NO:262)
--	---	-----------------

Source	Genbank Accession Number
<i>Shigella boydii</i> CDC 3083-94	YP_001881145.1
<i>Escherichia fergusonii</i> ATCC 35469	YP_002382013.1
<i>Salmonella enterica</i> subsp. <i>arizonae</i>	YP_001569590.1
<i>Citrobacter</i> sp. 30_2	ZP_04562837.1
<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578	YP_001336360.1
<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i> WPP14	ZP_03831287.1
<i>Enterobacter cancerogenus</i> ATCC 35316	ZP_03283474.1
<i>Pantoea</i> sp. At-9b	ZP_05730617.1
<i>Cronobacter turicensis</i>	CBA32510.1

FIG. 16I

<i>Dickeya dadantii</i> Ech586	ZP_05723897.1
<i>Erwinia tasmaniensis</i> Et1/99	YP_001907100.1
<i>Serratia proteamaculans</i> 568	YP_001479594.1
<i>Edwardsiella ictaluri</i> 93-146	YP_002934130.1
<i>Sodalis glossinidius</i> str. 'morsitans'	YP_455303.1
<i>Yersinia aldovae</i> ATCC 35236	ZP_04620215.1
<i>Providencia stuartii</i> ATCC 25827	ZP_02961167.1
<i>Photorhabdus asymbiotica</i>	YP_003040275.1
<i>Proteus mirabilis</i> HI4320	YP_002151524.1
<i>Candidatus Blochmannia</i> <i>pennsylvanicus</i> str. BPEN	YP_278005.1
<i>Glaciecola</i> sp. HTCC2999	ZP_03561088.1
<i>Vibrio cholerae</i> V51	ZP_04919940.1
<i>Wigglesworthia glossinidia</i> endosymbiont of <i>Glossina brevipalpis</i>	NP_871411.1
<i>Tolumonas auensis</i> DSM 9187	YP_002892770.1
<i>Actinobacillus</i> <i>pleuropneumoniae</i> serovar 1 str. 4074	ZP_00134992.2
<i>Aggregatibacter aphrophilus</i> NJ8700	YP_003007711.1
<i>Pseudoalteromonas</i> <i>tunicata</i> D2	ZP_01135065.1
<i>Vibriionales bacterium</i> SWAT-3	ZP_01816638.1

FIG. 16J

<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	NP_245276.1
<i>Mannheimia succiniciproducens</i> MBEL55E	YP_088783.1
<i>Haemophilus somnus</i> 129PT	YP_718877.1
<i>Shewanella loihica</i> PV-4	YP_001094535.1
<i>Aliivibrio salmonicida</i> LFI1238	YP_002262558.1

FIG. 17A

**AAR 7942 sequence**SEQ ID NO:195 - *Synechococcus elongatus* PCC7942 Synpcc7942 1594 DNA

```

1 atgttcggtc ttatcggta tctcaccagt ttggagcagg cccgcgacgt ttctcgagg
61 atgggtacg acgaatacgc cgatcaagga ttggagttt ggatgcgc tcctcctcaa
121 atcggtatg aaatcacatg caccatgtcc acaggcaagg tgatccacgg tegctacatc
181 gaatcggtt tcttgcggta aatgctggcg gcgcggcgct tcaaaacacgc cacgcgaaa
241 gttctcaatg ccattgtccca tgccccaaaa cacggcatcg acatctggc ctggggggc
301 ttatcctcgta ttatccatg gaatgtccatg ttggccaggat tgccgcaact gggcgacact
361 acttggatgttgaacgggtt caccacccgg aataactaca cggctacatg aatctgttaga
421 cagggtggaa cccgtgtctaa aacgcgtggc atcgacattt cccaagcgac atagcggtt
481 gtcggcgca ctggcgatata cggtagcgct gtctcgctt ggctcgaccc caaactgggt
541 gtcggtgatt tgatccgtac ggccggcaat caggagcgat tggataaccc gcaggctgaa
601 ctcggccggg gcaagattct gcocttgaa gcccgtctgc cggaaagctga ctttatcggt
661 tgggtggcca gtatcgctaa gggcgatgtt atcgacccag caaccctgaa gcaaccctgc
721 gtcctaatacg acggggggctaa ccccaaaaaac tggggcagaa aagtccaaagg tgagggcata
781 tatgtccatca atgggggggtt agttgaacat tgcttcgaca tcgactggca gatcatgtcc
841 gctgcagaga tggccggcccgagccatg atgtttcgctt gcttgcggaa ggcgtatgctc
901 ttggaatttg aaggctggca tactaacttccctggggcc gcaacccaaat cacgatcgag
961 aagatggaaag cgatcggtt ggcattcggtt cggccacggctt tccaaaccctt ggcattggca
1021 atttgaa

```

SEQ ID NO:196 - *Synechococcus elongatus* PCC7942 Synpcc7942 1594 protein (YP\_400611)

```

1 MFGLIGHLTS LEQARDVSRR MGYDEYADQG LEFWSSAPPQ IVDEITVTSA TGKVIHGRYI
61 ESCFLPEMLA ARRFKTATRK VLNAMSHAQK HGIDISALGG FTSIIFENFD LASLRQVRDT
121 TLEFERFTTG NHTHAYVICR QVEAAAKTLG IDITQATVAV VGATGDIGSA VCRWLDLKLG
181 VGLDLILTARN QERLDNLQAE LGRGKILPLE AALPEADFIV WVASMPQGVV IDPATLKQPC
241 VLIDGGYPKN LGSKVQGFGTI YVLNGGVVEH CFDIDWQIMLS AEMARPERQ MFACFAEAML
301 LEFEGWHTNF SWGRNQITIE KMEAIGEASV RHGFQPLALA I

```

**SEQ ID NO:213 – Nucleotide sequence of plasmid pCL-Ptrc-carB 'tesA**

```

CACTATACCAATTGAGATGGGCTAGTCATGATAATTACTAGTCCTTTCCCTTGAGTTGTTGGGTATCTGAAATTCTGCTAGACC
TTGCTGGAAAACCTGTAATTCTGCTAGACCCCTCTGTAATTCCGCTAGACCTTTGTGTGTTTTTTGTTATTCAGTGGT
TATAATTATAGAATAAAAGAAAGATAAAAAAGATAAAAAGATAAGATCCCAGCCCTGTGTATAACTCACTACTTTAGTCAGTT
CGCAGTATTACAAAAGGATGTCGCAAACGCTTGTCTCCTACAAAACACACCTTAAACCTAAAGGCgtcGGCATCCGCITA
CAGACAAGCTGTGACCGCTCCCGGAGCTGCACTGTGTCAGAGGTTTACCGTCATCACCGAAACGCCGAGGGCAGCAGATCAATT
CGCGCGCAAGGGCAAGCGCATGCAATTACGTTGACACCATGCAATGGTCAAAACCTTTCGGGTATGGCATGATACGCCCGG
AAGAGAGTCATTCAAGGTGGTGAATGTGAAACAGTAACGTTATACGNTCTCGCAGACTATGCCGTGCTCTTATCAGACCGTT
TCCCGCTGGTGAACCAAGGCCAGCCACGTTCTGCAAAACCCGGAAAAGTGGAAAGCCGATGGGGAGCTGAATTACATTCC
CAACCCGCTGGCACAACAACCTGGGGCAAACACTCGTTCGCTATTGGCGTGTGCCACCTCCAGTCAGGCCCTGCACGCCGTC
AAATTGTCGGCGGAIATTAAATCTCCCGCCGATCAACTGGTCCAGCGTGTGGTGTGATGGTGAAGACGAAGCGGGCTCGAAGGCC
TGTAAGCGCGGCTGCACAATCTCTCGCGCAACCGCTCAGTGGGCTGATCATTAACTACCGCTGGATGACCAGGATGCCATTG
TGTGGAAGCTGCTGCACTATGTTCCGGCGTTATTCTGTGATGTCCTCGCAGACACCCATCAACAGTATTATTTCTCCCATG
AAGACGGTACCGCAGCTGGCGTGGAGCATCTGGTCGCAATTGGCTCACCAGCAAATCGCGCTGTTAGCCCCCATTAAAGTTCTGTC
TCGGCGCTCTGCGTCTGGCTGGCATAAATCTCACTCGCAATCAAATTCAAGCGATAGCGGAACGGGAAGGCCACTGGAG
TGCCATGTCGGTTTCAACAAACCATGCAAATGCAATGCAATGAGGGCATCGTCCACTGCGATGCTGGTGTGCCAACGATCAGATGG
CGCTGGGCCAATGGGGCCATTACCGAGTCCGGCTCGCCGTTGGTGCAGGATATCTCGCTAGTGGGATACGACGATACCGAAGAC
AGCTCATGTTATATCCCGCCGTTAACCAACCATCAAACAGGATTTTCGGCTCTGGGGCAACCCAGCGTGGACCGCTTGTGCAACT
CTCTCAGGGCCAGGGCGTGAAGGGCAATCAGCTGTTGGCCGCTCACTGGTGAAGAAAACCCATGGCGCCAAATACGCAA
CCGCTCTCCCGCCGCTTGGCCGATTCAATTATGCACTGGCTGCAAGCTTACGCTGCTGCACTGGCTGCTGCAACG
AATTATGTAAGTAGCGCGAATTGATCTGGTTGACAGCTTATCATCGACTGCACGGTCACCAATGCTCTGGCGTCAAGGCAGC
CATCGGAAGCTGTGGTATGGCTGTGCAAGGTGTAATCACTGCAATAATTGCTGTCGCTCAAGGCAGCCTCCGTTCTGGATAATG
TTTTTGGCCGACATCATAACGGTTCTGGCAATATTCTGAATGAGCTGTGACAATTATCATCCGCTGTTAATGTGTGGA
ATTGTGAGCGGATAACAATTTCACACAGGAAACAGCGCCGCTGAGAAAAAGCGAAGCGGACTGCTCTTAAACAATTATCAGACA
ATCTGTGTPGGGCACTCGACCGGAATTATCGATTAACCTTATTAAAGAGGTATATTAATGTATCGATTAAGA

```

FIG. 17B

FIG. 17C

CATTCGCTCATGCCAGCCAGTCGGCGGCAGTCCATAGCGTTAAGGTTCAATTAGGCCCTAAATAGATCCTGTCAGGA  
 ACCGATCAAAGACTTCCCTCCCGCTGGACCTACCAAGGCACACGCTATGTTCTTGCTTCAGCAAGATAGCCAGATCAAT  
 GTCGATCGTGGCTGGCTCGAAGATAACCTGCAAGAATGTCATGCGCTGCCATTCTCCAATTGCAAGTTCCGCTTAGCTGGATAAC  
 GCCACGGAATGATGTCGTCGTGACAACAAATGCTGACTTCTACAGCGCGGAGAATCTCGCTCTCCAGGGAGCGGAAGTTCC  
 AAAAGCTCGTTGATCAAAGCTCCCGCGTTTCAAGGCCTAACCGTAACCAGCAAATCAATGACTGTCGGCT  
 CAGGCCGCATCCACTCGGGAGCGTACAAATGACGCCAGCAACGTCGGTCTGAGATGGCCTCGATGACGCCAACTACCTCTG  
 ATAGTTGAGTCGATACTCGGCGATCACCGCTCCCTCATGATGTTAACCTTCTTTAGGGCAGTGCCCTGCTGCGTAACATCG  
 TTGCTGCTCCATAACATCAAACATCGACCCACGGCGTAACCCGCTTCTGCTGTTGGATGCCGAGGCATAGACTGTACCCCCAAAAAA  
 ACAGCTATAACAACGCCATGAAAACGCCACTCGCCTGGCTTACACCCGCTTCTGGCTCGGCTCAAGGTTCTGGACCAAGTGGCTGAGCGCA  
 TACGCTACTTGCTCATACAGCTTACGAACGCCAGGCTATGTCACACTGGGCTGTGCTGCTTCCATCGTTTACCGGTGTCGCTCAC  
 CCGCACCTTGGCGAGCAGCGAAGTCGAGGCATTCTGCTCTGGCGAACAGCGCAAGCTTCCGCTCCACCGCATCGTCA  
 GGCATTTGGCGGCTTGCTGTTCTACGGCAAGGTCTGTCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGT  
 CGCGCGCTTGGCGGTGGTGCACCCGGATGAAGTGGTTCGCATCTCGGTTCTGGAAGGCAGCATGTTGTCGCCAG  
 CTTCGTATGGAACGGCAGCGCATGCTGAGGCTTGGCAACTGCGGGTCAAGGATCTGGATTTCTGATCACGGCACGATCATCGT  
 GCGGAGCGCAACGGCTCAAGGATCGGGCTTGATGTTACCGGAGAGCTGGCACCCAGCTGCGCAGCAGGGGAATTAAATTCC  
 CACGGTTTGTCTCCCGAACCGGCTTGCTGCTGTTAGTTATCAGAATCGCAGATCCGTTTCAAGCCGGTTGCTGCG  
 GCTGAAGCGCTATTCTCAGAATTGCCATGATTTTCCCACGGGAGGCTCACTGGCTCCGTGTTGCGCAGCTTGAT  
 TCGATAAGCAGCATCGCTGTTACGGCTGTCATGTCAGTGTGACTGTTGAGCTGTAACAAGTGTCTCAGGTGTTCAATTCTATGTT  
 TAGTTGCTTGTGTTACTGGTTACCTGTTCACTGTTGTTAGGTGTTACATGCTGTTCATCTGTTACATTGTCGATCTGTTCATGCTGAA  
 CAGTTTCAATGCAACGACAAAATCTGTAACACCTCTGATGATCTATCTATCTTACACCCTTCACTGTCATATGGACAGTTT  
 CCCATTGATATGTAACGGTAACAGCTGTCACCTTGTGTTAGTCTGATGCTGACTGATAGATAACAAGGCCATAAGAAC  
 CTCAGATCCTTCCGATTAGCAGTATGTTCTAGTGTGTTCTAGTGTGTTCTGTTCCGTTGAGCCATGAGAACGAAACCATGAGATCA  
 TACITACTTTGCTGACTCAAAATTTGCTCAAAACTGGTGAACCTGCAATTGCAAGTAAAGCATCGTGTAGTGTGTTCT  
 TAGTCGTTATGAGGATCTGATGTAATGGTTGTTGGTATTGTCACCAATTGTTATCTGGTTCTCAAGTCGG  
 TTACGAGATCATTGCTATCTAGTCACGGAAAATCAACGATACAGTCGGCGGCCCTCGTTATCAACCACCAATTTCATA  
 TTGCTAAGTGTAAATCTTACTTATGGTTCAAAACCCATTGGTAAGCCTTTAAACTCATCGTACTTATTTCAGCAT  
 TAACATGAACCTAAATTCAAGGCTAACATCTCATATTGCTCTGAGTTCTTCTGTTAGTTCTTTAATAACCACTCAT  
 AAATCTCTCATAGAGTATTGTTCAAAGACTAACATGTCAGATTATTTGTAATTTTAACTGGAAAAGATAAGGC  
 AATATCTTCACTAAAACCTAAATTCTAATTGCTTGAGAACTGGCATAGTTGTCACCTGGAAAATCTCAAAGCCTTAAAC  
 CAAAGGATTCTGATTCCACAGTCTCGTCATCAGCTCTGGTTGCTTAGCTAATACACCAAAAGCATTTCCACTGATGT  
 TCATCATCTGAGCGTATTGGTTAAAGTGAACGATACCGTCGTTCTTGTAGGGTTCAATCGTGGTTGAGTAGTGCC  
 ACACAGCATAAAATAGCTTGTGTTCATGTCGTTAACGACTAATCGCTAGTTGCTTGAACAAACTAAATT  
 AGACATACATCTAATTGGTCTAGGTGATTAAAT

TER Euglena gracilis (Q5EU90) without N-terminal transit peptide

## Amino acid sequence (SEQ ID NO:228)

```

1 ypqmaegfsg eatsawaaag pqwaaplva assalalwww aarrsvrrpl aalaelptav
 61 thlappmamf tttakviqpk irgfictth pigcekrvqe eiayarahpp tspgpkrvlv
121 igcstgygls tritaafgyq aatlgvflag pptkgrpaaa gwyntvafe kaaeaglyar
181 slngdafdst tkartveaik rdlgtvdllv ysiaapkrtd patgvlhkac lkpigatytn
241 rtvntdkaev tdvsiepas eeiadtvkvm ggedwelwiq alseagvlae gaktvaysyi
301 gpemtwpvyw sgtigeakk vekaakritq qygcpaypv akalvtqass aipvvplyic
361 llyrvmkek thegcieqmv rllttklype ngapivdeag rvrvddwema edvqqavkdl
421 wsqvstanlk disdfagyqt eflrlfgfqi dgdvdydqpv d veadlpsaaq q

```

fadA (YP026272)

## Nucleotide sequence (SEQ ID NO:229)

```

1 atggaacagg ttgtcattgt cgtgcattt cgccacccga tggccgttc gaaggcggt
 61 gctttcgta acgtgcgltc agaagatctc tccgcttatt taatgcgttag cctgtggcg
121 cgtaacccgg cgtggaagc ggccggccctc gacgatattt actgggttg tgcgcagcag
181 acgtggagc agggtttaa tatacgccgt aacgcggcgc tgctggcaga agtaccacac
241 tctgtcccg cggttaccgt taatcgcttgc tgtggttcat ccatgcaggc actgcgtac

```

## FIG. 17D

301 gcagcacgaa ttagatcatgac tggcgatgcg caggcatgtc tggttggcg cgtggagcat  
 361 atggccatg tgcccgatg tcacggcgatc gattttcacc ccggcccttag cccgaatgtc  
 421 gccaaagcg cgccatgtat gggcttaacg gcagaaatgc tggcgatgtat gcacggatgc  
 481 agccgtgaaa tgcaggatgc ctttgcccg cggtcacacg cccgcgcctg ggccgcac  
 541 cagtccggcc cattaaaaaa tgaatcatc cogacccgtg gtcacatgc cgcacggcgatc  
 601 ctgaagcgtt ttaattacga cgaatgtt gccccggaaa ccaccgttga agccctcgcc  
 661 acgctcgatc cggcgatgttgc tccagtaaac ggtatgttga eggggggcac atcttctgca  
 721 ctttcccgatg ggcgcgtgc catgtcggtg atgatgtaaa gccgcgccta tgaatttaggt  
 781 cttaaagccgc ggcgcgtgt ggcgtcgatg ggcgtcggtt gttgtgaccc atcgattatg  
 841 ggttacggcc cgggtccggc ctgcggaaactg ggcgtgaaaaa aagccccgtt ttctgccc  
 901 gatatacgcc tggttgaaaat gaacgaagcc tttgcggccg agatccgtcc atgtattaaa  
 961 gatctgggac taattqagca gattgacgag aagatcaacc tcaacgggtt cgcgcgtcg  
 1021 ctgggtcatc cgctgggttgc ttccgggtgc cgtatcagca ccacgcgttgc gaatctgtatg  
 1081 gaacgcggaaac acgttcagtt tggctggcg acgatgttgc tccggctggg tcagggtatt  
 1141 ggcgcgtgt ttgagccgggt ttaaa

Amino acid sequence (SEQ ID NO:230)

1 meqvvivdai rtpmqrskq afrmraedl sahlmrslslla rnpalceaaal ddiywqcvqq  
 61 tleqgfniar naallaevph svpavtvnrl cgssmqalhd aarmimtgdq qaclvggveh  
 121 mghvpmshgv dfhpqlsrnv akaagmmgl aemlarhmgi sremqdafaa rsharawaat  
 181 qsaafkneii ptqghdadgv lkqfnydevi rpettveala tlrpafdpvn gmvtagtssa  
 241 lsdgaaamlv msesrahelg lkprarvrsm avvgcdpsim gygpvpaskl alkkaglsas  
 301 digvfemnea faaqilpcik dlglieqide kinlnggaaia lghplgcsa risttlilm  
 361 erkdqvfgla tmciglgqgi atvferv

fadB (NP\_416843)

Nucleotide sequence (SEQ ID NO:231)

1 atgcgtttaca aaggcgacac cctgtacatt gactggctgg aagatggcat tgccgaactg  
 61 gtatgttgcgccccaggatc agttaataaa ctcgacactg cgaccgtcg cagcctcgcc  
 121 gaggccatcg gctgtcgatggaa acagcaatca gatctaaag ggctgtgtat ggcgttgc  
 181 aaagcggctt atatcgatgg tgcgtatc accgaatttt tgcgttgcgtt cctcgatcc  
 241 gaagaacagt taagtcgtg gctgcacttt gccaatagcg tgcgttgcgtt cctggaaagat  
 301 ctggccgtgc cgaccattgc tgcgtcaat ggctatgc tggccgttgc ctgcgtatgc  
 361 gtgcgttgcgca cccgttacgc tctggcgacg ccggatctgc gcatcggttgc gccggaaacc  
 421 aaactgggca tcatgcgttgc ctggccgttgc tctgtacgtt tgccacgtat gctggggcgat  
 481 gacagtgcgc tggaaatcat tgcgttgcgtt aaagatgtcg ggcgcgtatca ggcgttgc  
 541 atccgttgcgttgc tggatggcgat gtcgttgcgtt gacgttgcgtt gacgttgcgtt  
 601 ttacgccagg ccattaaacgg cgaccctcgac tggaaagcaa aacgtcagcc gaagctggaa  
 661 ccactaaaac tgagcaagat tgaagccacc atgagcttca ccacgtctaa aggatggtc  
 721 gcacaaaacag cggggaaaca ttatccggcc cccatcaccg cagtaaaaac cattgaagct  
 781 gccggccgtt ttggcgatgtt gacgttgcgtt aacatggaaa aacaaaatgtt tgcgttgcgt  
 841 ggcgcatacca acgaaagcccg cgcactggtc ggcattttcc ttaacgtatca atatgtaaaa  
 901 ggccaaagcgaa agaaactcacc caaagacgtt gaaaccccgaa aacaggccgc ggtgttgc  
 961 gcaggcatttca tcaacgacaa gtcgttaccc ctcggatgtt gacgttgcgtt gacgttgcgt  
 1021 atgaaaagatca tcaacgacaa gtcgttaccc ctcggatgtt gacgttgcgtt gacgttgcgt  
 1081 aacaaggcgc tttggcgccgg caagatcgat ggtctgtt gacgttgcgtt gacgttgcgt  
 1141 atccacccaa cgcgtcgacta cgcggattt gacccgttgcgtt gacgttgcgtt gacgttgcgt  
 1201 gttggaaaacc cggaaagtggaa aaaaggccgtt ctggcagaaa ccgaacaaaaa agtacggcc  
 1261 gataccgtgc tggcgatctaa cacttcaacc attccatataca ggcgttgcgtt gacgttgcgt

FIG. 17E

1321 gaacggcccgaaaacttctgcgggatgcacttcttaacccggtccaccgaatggcgttg  
1381 gtagaaattatcgcggcgaaaaagctcggacaaaccatcgcgaaagtgtgcctctgg  
1441 gcgagcaagaatggcaagacgcgattgtgttaacgcactccccggctcttttaaac  
1501 cgcgtgtgttccgtatttcgcgggttcagccagctgcgcgcacggcgcgatttc  
1561 cgcaagatcgacaaaagtgtatggaaaaacagttggctggccgatggggccggcatatctg  
1621 ctggacgttgcggcattgataccgcgcatacgctcaggctgtcatggcagcaggcttc  
1681 ccgcacggatgcagaaagaataccgcgcatacgacgcgtgtttga tgccaaacccgc  
1741 ttggtcagaagaacggcctcggttctggcgttataaaagaaagacagcaaagglaageccg  
1801 aaaaagaagaaagacggcgcgttgaaagactgtctggcagaaatggccgcgaatggcca  
1861 gattcagcgaaaagagattatcgccgcgtatgtatccgtgtttcaacgaaagccgc  
1921 cgcgtgttggatggaaaggcatatcgccactccggcggaaatcgatatcgatgttctac  
1981 ggctgggtttccctcggtccacggcgcgttccgcgtggctggacaccttcgttgc  
2041 gaaaaataccatcgatatgcgacagaatatacgcacacgcgttgcgttgc  
2101 gaaggctgcgtataaaagcgcgtataacgaaccgtactatcctccgttgagccagcc  
2161 cgtccggttqgcgacctgaaacggcttaa

Amino acid sequence (SEQ ID NO:232)

1 mlykgdtlyl dwledgjaiel vfdapgsvnk ldtatvaslg eaigvleqqc dlkglllrns  
61 kaafivgadi teflslflvp eeqlsqwlhf ansvfnrled lpvptiaavn gyalgggce  
121 vlatdyrlat pdlriglpet klgimpfggg svrmprmlga dsaleiiaag kdvgadqalk  
181 iglvvdgvvka eklvegakav lrqaingdld wkakrqpkle piklskieat msftiakgmv  
241 aqtagkhypa pitavktiea aarfgreeal nlenksfvpl ahtnearalv giflnqdqv  
301 gkakkltkdv ctpkqaavlg agimggijay qsawkgvpvv mkkdindkslt lgmtteaakll  
361 nkqlergkid glklagqvist ihptldyagf drvddivveav venpkvkav laeteqkv  
421 dtvlasnstt ipiselanal erpenfcgmh ffpnvhmpl veiirgekss detiakvvaw  
481 askmgktpiv vndcpqffvn rvlfpfyagf sqllrdgadf rkdkvmekq fgwpmgpay  
541 ldvvgidtah haqavmaagf pqrmqkdyrd aidalfdanr fgqknglqfw rykedskgkp  
601 kkeedaaved llaevsqpkd dfseeeeiar mmipmvnev rcleegijiat paeadmalvy  
661 glgfppfhgg afrwltdtgs akyldmaqgy qhlgpelyev eglrnkarhn epyyppvepa  
721 rpvqdlkta

**fadI** (NP\_416844)

Nucleotide sequence (SEQ ID NO:233)

```

  1 atgggtcagg ttttacccgtt ggttaccccg cagggcgcate gtatcgccat tgtagcggt
  61 ttacgtacgc cttttgcggcc tcaggcgacg gcttttcatg gcattcccg cgtttagttta
121 gggaaatgg tggtaggcga actgctggca cgcagcgaga tccccgcga agtgattgaa
181 caacttgtct ttggtcaggt cgtacaatag cctgaagccc ccaacattgc gcgtgaaatt
241 gttctcggtt cggaaatgaa tgtacatacc gatgcttaca gcgtcagcc cgcttcgcgt
301 accagttcc aggcatgtc aaacgtcgca gaaaggctga tggcggaaac tattcgagcg
361 gggattgcgg tgggggcaga ttcccttccg gtattgccaat tggcgtcag taaaaaaactg
421 gcgcgcgtgc tggttgatgt caacaaagct cgtaccatgt gccagcgact gaaactcttc
481 tctcgcttcg gtttgcgcga cttaatgccc gtaccacctg cggtagcaga atattctacc
541 ggtttgcggg tggggcgacac cgcagagcaa atggcgaaaaa cctacggcat caccggagaa
601 cagaagatg cattagcgca cgggtcgcat cagcggtccg ctcaggcatg gtcagacggg
661 aaactcaaag aagagggtat gactgccttt atcccttcctt ataaacaacc gcttgcgaa
721 gacaacaata ttgcgggtaa ttccctcgat gecgattacg caaagctcg gcccggqitt
781 gatcgcaaac acggaaacggt aacggcgca aacagtacgc cgctgaccega tggcggcgca
841 ggggtgtatcc tggatgactga atccccggcg aaaaatttag ggctgggtcc gctgggttat
901 ctgcgcacgt acggatcac tggatgtat gctggcagg acatgttgct cggttccagcc
961 tggtaacac cgttggcgct ggagcggtcc gtttgcacga tgagcgatct gacattgatc
1021 gatatgcacg aaggcttgc agtgcagacg ctggcgaata ttcaatgtct gggtagtgaa
1081 cgttttgctc gtgaagcaat gggggcggtca catgccactg gcgaaatgttca cgatagcaaa

```

## FIG. 17F

1141 tttaacgtgc ttggcggttc gattgcttac gggcattccct tgcggcgac cggcgcgccg  
 1201 atgattaccc agacattgca tgaacttgc cgtcgccgc gtggatttgg tttagttacc  
 1261 gcctgtgtcg ccggtgccc tggcgccga atggtttgg aggccgaaata a

## Amino acid sequence (SEQ ID NO:234)

1 mgqlplvtr qgdriaiivsg lrtpfarqat afhgipavdl gkmvvgella rseipaevie  
 61 qlvfgqvqvqm peapniarei vlgtnmnvt daysvsraca tsfqavanya eslmagtira  
 121 giaggadsss vlpigvskl arvlvdvnka rtmsqr1klf srlrlrdlmp vppavaeyst  
 181 glrmrgdtaeq maktygitre qqdalahrsh qraaqawsdg klkeevmtaf ippykplve  
 241 dnnirgnssl adyaklrpaf drkhgtvtaa nstpltdgaa avilmtesra kelglvplgy  
 301 lrsyaftaid vwqdmllgpa wstplalera gltmsdltli dmheafaqt laniqligse  
 361 rfarealgra hatgevddsk fnvlggsiay ghpfaatgar mitqtlhelr rrgggfglvt  
 421 acaaqqlgaa mvleae

fadJ (NP\_416843)

## Nucleotide sequence (SEQ ID NO:235)

1 atggaaatga catcagcggtt tacccttaat gttcgtcttgg acaacattgc cgtttatcacc  
 61 atcgcacgtac cgggtgagaa aatgaatacc ctgaaggcgg agtttgcctc gcaggtgcgc  
 121 gccattatta agcaactccg tgaaaacaaa gagttgcggag gctgtgggtt tgcgtccgct  
 181 aaaccggaca acttcattgc tggcgcagac atcaacatga tggcaactg caaaacggcg  
 241 caagaagcgg aagctctggc gccgcaggcc caacagttga tggcggagat tcattgtttg  
 301 cccattcagg ttatcgcggc tattcatggc gcttgcctgg gtgggtggct ggagttggcg  
 361 ctggcgtgcc acggtcgcgt ttgtactgac gatcctaaaaa cggtgctcgg tttgcctgaa  
 421 gtacaacttgc gattgttacc cgggttcaggc ggcacccaggc gtttacccggc tctgtataggc  
 481 gtcagcacag cattagagat gatcctcacc gaaaaacaaac ttccggcgaa acaggcattaa  
 541 aagctggggc tggtgatga cgttgcggc cactccatttc tgctggaaaggc cgctgtttag  
 601 ctggcaaaga aggagcgcacc atctcccgcc cctctacccgt tacgcggcggc tattctggcg  
 661 gggccgttag gtcgtgcgtc gctgtccaaa atggcggca agaaaaacaga acacaaaaact  
 721 caaggcaatt atccggcgc acagacgcattt ctggagggtt tgaaacggg attagcgcag  
 781 ggcaccagca gcggttatga cgcggaaatc cggggcgtttt gogaactggc gatgacgcac  
 841 caatcgcagg cgctgtgtatc tatctttttt gccagttacgg acgtgaagaa agatccccgc  
 901 agtgcgtgcgc cgcctgcgc attaacacgc gtggggattt tagtgtgtgg ctgtatggc  
 961 ggcggattttt cttatgtcac tgctgtaaa ggggggattt cggcgtccaaat taaagatata  
 1021 aaccgcagg gcataaatca tgcgtgaag tacagttggg atcagctgga gggcaagtt  
 1081 cgcgcgtgc atctcaaaggc cagcgaaacgt gacaaacaggc tggcattaaat ctccggaaacg  
 1141 acggactatc ggggtttgc ccattcgat ctgattattt aagcggtgtt tgaaaatctc  
 1201 gaattgaaac aacagatggt ggcggaaatggt gagcaaaattt ggcgcgtca taccatctt  
 1261 gcttcgataa cgtcatctt accgattggt gatatcgcc ctcacccac gcgcacgtttag  
 1321 caagtgtatcg gcttcgttccg gtggaaaaaa tgccgtgtt ggagattatt  
 1381 ccttcgtccgg ggacatcgcc gcaaaaccatc gtcaccacag taaaactggc gaaaaaaacag  
 1441 ggtaaaacgc caattgtgtt gctgtggccaa gccgggtttt acgtcaatcg catcttagcg  
 1501 ctttacattt atgaagatcat cccatgttgc accccaagggtt aacgggttgc gcacattgtat  
 1561 ggcgcgttag tggaaattttt ttttccgtt ggcggccatcc aacttttggc tgaggtagga  
 1621 atcgacacccg ggactaaaaat tatttcgttgc tggaaaggccg cttatggaga acgttttagc  
 1681 ggcgcgtccaa atgttgcgttcc ttcaattttt aacgcacgttgc gcaaaaggcgg aaaaaatggc  
 1741 cgggggttttcc atctttatgg tcagaaaggccg ctgttttttca aaaaacagggtt cgatccccgc  
 1801 atttacccgc tgattggccac acaaggccg gggccatcc acggcaccgc ggttgcgttgg  
 1861 cgggtgtgttgc tggatgtatggt gatgtggccatcc acgttttttggc tgaggtagga  
 1921 agcgtgcgttgc acggggatattt tggcgccgttgc tttggcattt gttttccgc atttctcggt  
 1981 ggaccgttcc gctatatcgat ttcttcgttgc gccggccatcc tggatgttgc aatgtcaacgc  
 2041 cttggccacgc agtatttttcc cgttttacc ctttcgttgc gtttgcgttgc gatggggcgc  
 2101 cgtggggaaa gtttttggaa aacaactgc aactgacccgc aataaa

**FIG. 17G**

Amino Acid sequence (SEQ ID NO:236)

```

1 memtsaftln vrldniav it idvpgekmnt lkaefasqvr aiikqlrenk elrgvvfvs
61 kpdnfiagad inmignckta qeaealarqq qqlmaeihal piqvaaiah aclgglela
121 lachgrvctd dpktvlgipe vqlglpgsg gtqrlprlig vstalemit gkqlrakqal
181 klglvddvvp hsilleaave lakkerpssr plpvrerila gplgrallfk mvgkktchkt
241 qgnypateri levvetglaq gtssgydaea rafgelamtp qsgalrsiff astdvkkdpg
301 sdappaplns vg ilggglmg ggiaytack agipvrikdi npqginhalk yswdqlegkv
361 rrrhlkaser dkqlalisgt tdyrgfahrd lieavfenl elkqqmvaev eqncaahitf
421 asntsslpig diaahatrpe qvighffsp vekmplveii phagtsaqt attvklakkq
481 gktpivvrdk agfyvnrla pyineairml tqgervehid aalvkfgfpv gpiqlldenv
541 idtgtkiipv leaaygerfs apanvssil nddrkgrkng rgfyllygqkg rkskqvdpa
601 iypligtqgg grisapvae rcvmlnea vrcvdeqvir svrdgdigav ffigfppflg
661 gpfryidslg agevvaimqr latqygsrft pcerl vemga rgesfwkta tdlq

```

**fabI (NP 415804)**

Nucleotide sequence (SEQ ID NO:237)

```

1 atgggttttc ttcccgtaa ggcattctg gtaaccggtg ttgccagcaa actatccatc
61 gcctacggta tcgctcaggc gatgcaccgc gaaggagctg aactgcatt cacctaccag
121 aacgacaaac taaaaggccg cgtagaagaa tttgcccctc aattgggtc tgacatcggtt
181 ctgcagtgcg atgttgcaga agatgccagc atcgacacca tttcgctga actggggaaa
241 qtttggccga aatttgcacgg tttcgatcac tctattgggt ttgcacctgg cgatcaqctg
301 gatggtgact atgttaacgc cggttaccgtt gaaggcttca aaattggccca cgacatcagc
361 tcctacagct tcggttgcataa ggcaaaagct tgccgctcca tgctaatcc gggttctggc
421 ctgctgaccc ttccctaccc tggcgcttag cgccgttatcc cgaactacaa cgttatgggt
481 ctggcaaaag cgtctctgga agcaacgtg cgctatatgg cgaacgcgtt ggtccggaa
541 gytgtcgltg ttaacccat ctctgttgtt ccgtatccgtt ctctgcggc ctccggatc
601 aaagacttcc gcaaaatgtt ggctcattgc gaagccgtt ccccgattcg ccgtaccgtt
661 actattgaag atgtggtaa ctctgcggca ttccctgtgtt ccgtatcttc tgccggatc
721 tccggtaag tggtccacgt tgacggcggt ttcatcgattt ctgcaatgaa cgaactcgaa
781 ctgaaataa

```

Amino acid sequence (SEQ ID NO:238)

```

1 mgflsgkril vtgvasklsi aygiaqamhr egaelaftyq ndklkgrvce faaqlgsdiv
61 lqcdvaedas idtmfaelgk vwpkfdgfvh sigfapgdql dgdvnavtr egfkiahdis
121 sysfvamaka crsmlnpgsa lltlsylgae raipnynvnmg lakasleanv rymanamgpe
181 gvrvnaisag pirtlaasi kdfkmlahc eavtpirrtv tiedvgnssaa flcsdlsagi
241 sgevvhvdgg fsiaamnele lk

```

**tesB (NP 414986)**

Nucleotide sequence (SEQ ID NO:239)

```

1 atgagtcagg cgctaaaaaa ttactgaca ttgttaaattc tggaaaaaaat tgaggaagga
61 ctctttcgcc gcccaggatga agattttagt ttacggccagg tggggccgg ccaggcgtt
121 gtcaggccct tggatgtgc aaaagagacc gtccctgaaag agcggtgtt acattcggtt
181 cacagctact ttcttcgccc tggcgatagt aagaagccga ttatttatga tggaaaccc
241 ctgcgtgacg gtaacagctt cagcccccgc cgggttgcgtt ctattaaaa cggcaaaccg
301 attttttata tgactgcctc ttccaggca ccagaagccg gtttgcaca tcaaaaaaca
361 atggccgtcc cgccaggccc tgatggccct ccttcggaaa ccaatcgcc ccaatcgctt
421 ggcacactgc tgccggcagt gctgaaagat aaattcatct ggcacgttcc gctggaaagtc
481 cgtccggatq agtttccatcc cccactgaaa ggtcacgtcg cagaacccaca tggcagggt
541 tggatccgcg caaatggtag cgtccggat gacctgcgcg ttcatcgat tctgcgttgcgtt
601 tacgcttcgtt atcttaactt cctgcggta gctctacagc cgcacggcat cggtttttc

```

**FIG. 17H**

661 gaaccgggga ttccagattgc caccattgac cattccatgt ggttccatcg cccgtttaat  
721 ttgaatgaat ggctgcgtga tagcgtggag agcacctcgg cgtccagcgc acgtggcttt  
781 gtgcgcggtg agtttatac ccaagacggc gtactggttg cctcgaccgt tcaggaaggg  
841 gtgatgcgtatcacaatta a

## Amino acid sequence (SEQ ID NO:240)

1 msqalknllt llnekiaeeg lfrqgqsedlg lrqvfgqqvv gqalyaaket vpeerlvhsf  
61 hsyflrpgds kkpiiydvet lrdgnfsar rvaaiqnqkp ifymtasfqa peagfehqkt  
121 mpsapapdgl psetqiaqsl ahllppvlkd kficdrplev rpvefhnplk ghvaephrqv  
181 wirangsvpd dlrvhqyllum yasdlnflpv alqphgigfl epgiqiatiid hsmwfhrpfn  
241 lnewlllysve stsassargf vrgefytdqdg vlvastvqeg vmrnhn

Acrl Acinetobacter (YP 047869)

## Nucleotide sequence (SEQ ID NO:241)

1 TTGATATCAA TCAGGGAAAA ACGCGTGAAC AAAAAACTTG AAGCTCTCTT CCGAGAGAAAT  
61 GTAAAAGGTAA AAGTGGCTTT GATCACTGGT GCATCTAGTG GAATCGGTTT GACGATTGCA  
121 AAAAGAATTG CTGCGGCAGG TGCTCATGTA TTATTGGTTG CCCGAACCCA AGAAACACTG  
181 GAAGAAGTGA AAGCTGCAAT TGAACAGCAA GGGGGACAGG CCTCTATTTT TCCTTCTGAC  
241 CTGACTGACA TGAATGCGAT TGACCAGTTA TCACAAACAAA TTATGGCCAG TGTCGATCAT  
301 GTCGATTTC TGATCAATAA TGCAGGGCGT TCGATTGCC CGGCCGTACA CGAGTCGTTT  
361 GATCGCTTCC ATGATTGTTGA ACCGACCATG CAGCTGAATT ACTTTGGTGC GGTACGTTTA  
421 GTGTTAAATT TACTGCCACA TATGATTAAG CGTAAAAATG GCCAGATCAT CAATATCAGC  
481 TCTATTGGTG TATTGGCCAA TGCAGACCGT TTTTCTGCTT ATGTCGCGTC TAAAGCTGCG  
541 CTGGATGCCT TCAGTCGCTG TCTTTCAGCC GAGGTACTCA AGCATAAAAT CTCATTAC  
601 TCGATTATA TGCCATTGGT GCGTACCCCA ATGATCGCAC CCACCAAAAT TTATAAATAC  
661 GTGCCACCGC TTCCCCAGA AGAAGCCGCA GATCTCATTG TCTACGCCAT TGTGAAACGT  
721 CCAAAACGTA TTGCGACGCA CTTGGGTCTGT CTGGCGTCAA TTACCTATGC CATCGCACCA  
781 GACATCAATA ATATTCTGAT GTCGATTGGA TTTAACCTAT TCCCAAGCTC AACGGCTGCA  
841 CTGGGTGAAC AGGAAAAATT GAATCTGCTA CAACGTGCCT ATGCCCGCTT GTTCCCAGGC  
901 GAACACTGGT AA

## Amino acid sequence (SEQ ID NO:242)

1 misirekrvn kklealfren vkgkvalitg assgigltia kriaaagahv llvartqetl  
61 eevkaiaeegg ggqasifpcd ltdmnaidql sqqimasvdh vdflinnagr sirravhesf  
121 drfhdfertm qlnyfgavr lvnllphmik rkngqiinis sigylanatr fsayvaskaa  
181 ldafsrclsa evlkhkisit siymplvrt miaptkiyky vptlspeea dlivyaivkr  
241 pkriathlgr lasityaiap dinnilmsig fnlfpsstaa lgeqeklnll qrayarlfpq  
301 ehw

## 1

**METHODS AND COMPOSITIONS FOR PRODUCING FATTY ALCOHOLS****CROSS-REFERENCE TO RELATED APPLICATIONS**

This application is a continuation of copending U.S. patent application Ser. No. 12/575,430, filed Oct. 7, 2009, which application claims the benefit of U.S. Provisional Application No. 61/109,131, filed Oct. 28, 2008, the contents of all of which are herein expressly incorporated by reference in their entirety.

**INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ELECTRONICALLY**

Incorporated by reference in its entirety herein is a computer-readable nucleotide/amino acid sequence listing submitted concurrently herewith and identified as follows: One 968,275 byte ASCII (Text) file named "LS00016 PCT\_SeqL-stg in Txt" created on Mar. 30, 2011.

**BACKGROUND OF THE INVENTION**

Petroleum is a limited, natural resource found in the Earth in liquid, gaseous, or solid forms. Petroleum is primarily composed of hydrocarbons, which are comprised mainly of carbon and hydrogen. It also contains significant amounts of other elements, such as, nitrogen, oxygen, or sulfur, in different forms.

Petroleum is a valuable resource, but petroleum products are developed at considerable costs, both financial and environmental. First, sources of petroleum must be discovered. Petroleum exploration is an expensive and risky venture. The cost of exploring deep water wells can exceed \$100 million. Moreover, there is no guarantee that these wells will contain petroleum. It is estimated that only 40% of drilled wells lead to productive wells generating commercial hydrocarbons. In addition to the economic cost, petroleum exploration carries a high environmental cost. For example, offshore exploration disturbs the surrounding marine environments.

After a productive well is discovered, the petroleum must be extracted from the Earth at great expense. During primary recovery, the natural pressure underground is sufficient to extract about 20% of the petroleum in the well. As this natural pressure falls, secondary recovery methods are employed, if economical. Generally, secondary recovery involves increasing the well's pressure by, for example, water injection, natural gas injection, or gas lift. Using secondary recovery methods, an additional 5% to 15% of petroleum is recovered. Once secondary recovery methods are exhausted, tertiary recovery methods can be used, if economical. Tertiary methods involve reducing the viscosity of the petroleum to make it easier to extract. Using tertiary recovery methods, an additional 5% to 15% of petroleum is recovered. Hence, even under the best circumstances, only 50% of the petroleum in a well can be extracted. Petroleum extraction also carries an environmental cost. For example, petroleum extraction can result in large seepages of petroleum rising to the surface. Moreover, offshore drilling involves dredging the seabed which disrupts or destroys the surrounding marine environment.

Since petroleum deposits are not found uniformly throughout the Earth, petroleum must be transported over great distances from petroleum producing regions to petroleum consuming regions. In addition to the shipping costs, there is also the environmental risk of devastating oil spills.

## 2

In its natural form, crude petroleum extracted from the Earth has few commercial uses. It is a mixture of hydrocarbons (e.g., paraffins (or alkanes), olefins (or alkenes), alkynes, napthenes (or cycloalkanes), aliphatic compounds, aromatic compounds, etc.) of varying length and complexity. In addition, crude petroleum contains other organic compounds (e.g., organic compounds containing nitrogen, oxygen, sulfur, etc.) and impurities (e.g., sulfur, salt, acid, metals, etc.).

Hence, crude petroleum must be refined and purified before it can be used commercially. Due to its high energy density and its easy transportability, most petroleum is refined into fuels, such as transportation fuels (e.g., gasoline, diesel, aviation fuel, etc.), heating oil, liquefied petroleum gas, etc.

Crude petroleum is also a primary source of raw materials for producing petrochemicals. The two main classes of raw materials derived from petroleum are short chain olefins (e.g., ethylene and propylene) and aromatics (e.g., benzene and xylene isomers). These raw materials are derived from longer chain hydrocarbons in crude petroleum by cracking it at considerable expense using a variety of methods, such as catalytic cracking, steam cracking, or catalytic reforming. These raw materials are used to make petrochemicals, which cannot be directly refined from crude petroleum, such as monomers, solvents, detergents, or adhesives.

One example of a raw material derived from crude petroleum is ethylene. Ethylene is used to produce petrochemicals, such as polyethylene, ethanol, ethylene oxide, ethylene glycol, polyester, glycol ether, ethoxylate, vinyl acetate, 1,2-dichloroethane, trichloroethylene, tetrachloroethylene, vinyl chloride, and polyvinyl chloride. An additional example of a raw material is propylene, which is used to produce isopropyl alcohol, acrylonitrile, polypropylene, propylene oxide, propylene glycol, glycol ethers, butylene, isobutylene, 1,3-butadiene, synthetic elastomers, polyolefins, alpha-olefins, fatty alcohols, acrylic acid, acrylic polymers, allyl chloride, epichlorohydrin, and epoxy resins.

These petrochemicals can then be used to make specialty chemicals, such as plastics, resins, fibers, elastomers, pharmaceuticals, lubricants, or gels. Particular specialty chemicals that can be produced from petrochemical raw materials are fatty acids, hydrocarbons (e.g., long chain, branched chain, saturated, unsaturated, etc.), fatty alcohols, esters, fatty aldehydes, ketones, lubricants, etc.

Fatty alcohols have many commercial uses. Worldwide annual sales of fatty alcohols and their derivatives are in excess of US\$1 billion. The shorter chain fatty alcohols are used in the cosmetic and food industries as emulsifiers, emollients, and thickeners. Due to their amphiphilic nature, fatty alcohols behave as nonionic surfactants, which are useful in personal care and household products, for example, detergents. In addition, fatty alcohols are used in waxes, gums, resins, pharmaceutical salves and lotions, lubricating oil additives, textile antistatic and finishing agents, plasticizers, cosmetics, industrial solvents, and solvents for fats.

Aldehydes are used to produce many specialty chemicals. For example, aldehydes are used to produce polymers, resins (e.g., Bakelite), dyes, flavorings, plasticizers, perfumes, pharmaceuticals, and other chemicals. Some are used as solvents, preservatives, or disinfectants. Some natural and synthetic compounds, such as vitamins and hormones, are aldehydes. In addition, many sugars contain aldehyde groups.

Obtaining these specialty chemicals from crude petroleum requires a significant financial investment as well as a great deal of energy. It is also an inefficient process because frequently the long chain hydrocarbons in crude petroleum are

cracked to produce smaller monomers. These monomers are then used as the raw material to manufacture the more complex specialty chemicals.

In addition to the problems with exploring, extracting, transporting, and refining petroleum, petroleum is a limited and dwindling resource. One estimate of world petroleum consumption is 30 billion barrels per year. By some estimates, it is predicted that at current production levels, the world's petroleum reserves could be depleted before the year 2050.

Finally, the burning of petroleum based fuels releases greenhouse gases (e.g., carbon dioxide) and other forms of air pollution (e.g., carbon monoxide, sulfur dioxide, etc.). As the world's demand for fuel increases, the emission of greenhouse gases and other forms of air pollution also increases. The accumulation of greenhouse gases in the atmosphere can lead to an increase global warming. Hence, in addition to damaging the environment locally (e.g., oil spills, dredging of marine environments, etc.), burning petroleum also damages the environment globally.

Due to the inherent challenges posed by petroleum, there is a need for a renewable petroleum source that does not need to be explored, extracted, transported over long distances, or substantially refined like petroleum. There is also a need for a renewable petroleum source which can be produced economically without creating the type of environmental damage produced by the petroleum industry and the burning of petroleum based fuels. For similar reasons, there is also a need for a renewable source of chemicals which are typically derived from petroleum.

One method of producing renewable petroleum is by engineering microorganisms to produce renewable petroleum products. Some microorganisms have a natural ability to produce chemicals. For example, yeast has been used for centuries to produce ethanol (e.g., beer, wine, etc.). In recent years, through the development of advanced biotechnologies, it is possible to metabolically engineer an organism to produce bioproducts that were never previously produced. Products, such as chemicals, derived from these cellular activities are known as bioproducts. Fuels produced these cellular activities are known as biofuels. Biofuels are a renewable alternative fuel to petroleum based fuels. Biofuels can be substituted for any petroleum based fuel (e.g., gasoline, diesel, aviation fuel, heating oil, etc.). Biofuels can be derived from renewable sources, such as plant matter, animal matter, or even waste products. These renewable sources are collectively known as biomass. One advantage of biofuels over petroleum based fuels is that they do not require expensive and risky exploration or extraction. In addition, biofuels can be locally produced. Hence, they do not require transportation over long distances. Moreover, biofuels can be made directly without the need for expensive and energy intensive refining as is needed with refining crude petroleum. In other circumstances, the biofuel may require a limited and cost-effective level of refining. Furthermore, the use of biofuels improves the environment by reducing the amount of environmentally harmful emissions (e.g., green house gases, air pollution, etc.) released during combustion. For example, biofuels maintain a balanced carbon cycle because biofuels are produced from biomass, a renewable, natural resource. While the burning of biofuels will release carbon (e.g., as carbon dioxide), this carbon will be recycled during the production of biomass (e.g., the cultivation of crops), thereby balancing the carbon cycle unlike petroleum based fuels.

For similar reasons, biologically derived chemicals offer the same advantages as biofuels over petroleum based fuels. Biologically derived chemicals are a renewable alternative to petrochemicals. Biologically derived chemicals, such as

hydrocarbons (e.g., alkanes, alkenes, or alkynes), fatty alcohols, esters, fatty acids, fatty aldehydes, and ketones are superior to petrochemicals because they are produced directly without extensive refining. Unlike petrochemicals, biologically derived chemicals do not need to be refined like crude petroleum to recover raw materials which must then be further processed to make more complex petrochemicals. Biologically derived chemicals are directly converted from biomass to the desired chemical product.

10

## SUMMARY OF THE INVENTION

The invention is based, at least in part, on the identification of genes that encode fatty aldehyde biosynthetic polypeptides and fatty alcohol biosynthetic polypeptides, which can be used to produce fatty aldehydes that can subsequently be converted into fatty alcohols. Accordingly, in one aspect, the invention features a method of making a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof. In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272 with one or more amino acid substitutions, additions, insertions, or deletions, and the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endog-

enous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at

least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272. In some embodiments, the amino acid sequence is the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16.

In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271, or to a fragment thereof.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino

acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polynucleotide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method comprises expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising the amino acid of SEQ ID NO:16, or a variant thereof. In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the polypeptide comprises the amino acid sequence of SEQ ID NO:16 with one or more amino acid substitutions, additions, insertions, or deletions,

wherein the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such

as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising an amino acid sequence having at least about 70% sequence identity to the amino acid sequence of SEQ ID NO:16.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell.

In some embodiments, the amino acid sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:16. In some embodiments, the amino acid sequence is SEQ ID NO:16.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/

## 11

isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:15, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of

## 12

fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:15, or to a fragment thereof.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty aldehyde biosynthetic nucleotide sequence having at least about 70% sequence identity to a nucleotide sequence listed in FIG. 8. In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271. In some embodiments, the nucleotide sequence is SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 263, 265, 267, 269, or 271.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

In some embodiments, the recombinant vector is a plasmid.

13

In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty aldehyde biosynthetic nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:15.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments,

14

the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

10 In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:15. In some embodiments, the nucleotide sequence is the nucleotide sequence of SEQ ID NO:15.

15 In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

20 In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

25 In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

30 In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

## 15

In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

In some embodiments, the recombinant vector is a plasmid.

In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty aldehyde biosynthetic polypeptide comprising (i) SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10; (ii) SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; and/or (iii) SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11; wherein the polypeptide has carboxylic acid reductase activity. In some embodiments, the polypeptide has fatty acid reductase activity.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the polypeptide is about 1,000 amino acids to about 2,000 amino acids in length. In certain embodiments, the polypeptide is about 1,000 amino acids in length, about 1,050 amino acids in length, about 1,100 amino acids in length, about 1,150 amino acids in length, about 1,200 amino acids in length, about 1,250 amino acids in length, about 1,300 amino acids in length, about 1,400 amino acids in length, about 1,500 amino acids in length, about 1,600 amino acids in length, about 1,700 amino acids in length, about 1,800 amino acids in length, about 1,900 amino acids in length, or about 2,000 amino acids in length. In other embodiments, the polypeptide is up to about 2,000 amino acids in length, up to about 1,900 amino acids in length, up to about 1,800 amino acids in length, up to about 1,700 amino acids in length, up to about 1,600 amino acids in length, up to about 1,500 amino acids in length, up to about 1,400 amino acids in length, up to about 1,300 amino acids in length, up to about 1,250 amino acids in length, up to about 1,200 amino acids in length, up to about 1,150 amino acids in length, up to about 1,100 amino acids in length, up to about 1,050 amino acids in length, or up to about 1,000 amino acids in length. In

## 16

other embodiments, the polypeptide is more than about 1,000 amino acids in length, more than about 1,050 amino acids in length, more than about 1,100 amino acids in length, more than about 1,150 amino acids in length, more than about 1,200 amino acids in length, more than about 1,250 amino acids in length, more than about 1,300 amino acids in length, more than about 1,400 amino acids in length, more than about 1,500 amino acids in length, more than about 1,600 amino acids in length, more than about 1,700 amino acids in length, more than about 1,800 amino acids in length, more than about 1,900 amino acids in length, or about 2,000 amino acids in length.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty alcohol biosynthetic polypeptide in the host cell. In particular embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of making a fatty alcohol. The method includes expressing in a host

cell a gene encoding a fatty alcohol biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof. In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194 with one or more amino acid substitutions, additions, insertions, or deletions, and the polypeptide has alcohol dehydrogenase activity.

In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has alcohol dehydrogenase activity.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a bacterium. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty alcohol biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a gene encoding a fatty alcohol biosynthetic polypeptide comprising an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194. In some embodiments, the amino acid sequence is SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074; fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polypeptide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty alcohol biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a polynucleotide that hybridizes to a complement of the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or to a fragment thereof, wherein the polynucleotide encodes a polypeptide having alcohol dehydrogenase activity.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the polynucleotide hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions, to a complement of the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or to a fragment thereof.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the polynucleotide is from a bacterium, a plant, an insect, a yeast, a fungus, or a mammal.

In certain embodiments, the polypeptide is from a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, bacterial cell, or any other organism described herein. In some embodiments, the bacterium is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*. In other embodiments, the bacterium is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola*, or *Clavibacter michiganensis*.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for the fatty aldehyde biosynthetic polypeptide.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes expressing in a host cell a recombinant vector comprising a fatty alcohol biosynthetic nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence is of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193.

In some embodiments, the method further includes isolating the fatty alcohol from the host cell. In some embodiments, the fatty alcohol is present in the extracellular environment. In certain embodiments, the fatty alcohol is isolated from the extracellular environment of the host cell. In some embodiments, the fatty alcohol is secreted from the host cell. In alternative embodiments, the fatty alcohol is transported into the extracellular environment. In other embodiments, the fatty alcohol is passively transported into the extracellular environment.

In some embodiments, the recombinant vector further comprises a promoter operably linked to the nucleotide

sequence. In certain embodiments, the promoter is a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

In other embodiments, the recombinant vector comprises at least one sequence selected from the group consisting of (a) a regulatory sequence operatively coupled to the nucleotide sequence; (b) a selection marker operatively coupled to the nucleotide sequence; (c) a marker sequence operatively coupled to the nucleotide sequence; (d) a purification moiety operatively coupled to the nucleotide sequence; (e) a secretion sequence operatively coupled to the nucleotide sequence; and (f) a targeting sequence operatively coupled to the nucleotide sequence.

In some embodiments, the recombinant vector is a plasmid.

In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector. In some embodiments, the nucleotide sequence is stably incorporated into the genomic DNA of the host cell, and the expression of the nucleotide sequence is under the control of a regulated promoter region.

In some embodiments, the method further includes modifying the expression of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, modifying the expression of a gene encoding a fatty acid synthase includes expressing a gene encoding a fatty acid synthase in the host cell and/or increasing the expression or activity of an endogenous fatty acid synthase in the host cell. In alternate embodiments, modifying the expression of a gene encoding a fatty acid synthase includes attenuating a gene encoding a fatty acid synthase in the host cell and/or decreasing the expression or activity of an endogenous fatty acid synthase in the host cell. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In some embodiments, the method further includes expressing a gene encoding a fatty aldehyde biosynthetic polypeptide in the host cell. In particular embodiments, the fatty aldehyde biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 264, 266, 268, 270, or 272, or a variant thereof.

In other embodiments, the host cell is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type host cell. In some embodiments, the host cell is genetically engineered to express an attenuated level of an acyl-CoA synthase relative to a wild type host cell. In particular embodiments, the host cell expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the genetically engineered host cell comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the host cell is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the host cell comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the host cell is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the host cell comprises a knockout of

fabB or a gene listed in FIG. 16. In yet other embodiments, the host cell is genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the method further includes culturing the host cell in the presence of at least one biological substrate for a fatty alcohol biosynthetic polypeptide.

In any of the aspects of the invention described herein, the host cell can be selected from the group consisting of a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, and bacterial cell.

In some embodiments, the host cell is a Gram-positive bacterial cell. In other embodiments, the host cell is a Gram-negative bacterial cell.

In some embodiments, the host cell is selected from the genus *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*.

In certain embodiments, the host cell is a *Bacillus lenthus* cell, a *Bacillus brevis* cell, a *Bacillus stearothermophilus* cell, a *Bacillus licheniformis* cell, a *Bacillus alkalophilus* cell, a *Bacillus coagulans* cell, a *Bacillus circulans* cell, a *Bacillus pumilus* cell, a *Bacillus thuringiensis* cell, a *Bacillus clausii* cell, a *Bacillus inegaterium* cell, a *Bacillus subtilis* cell, or a *Bacillus amyloliquefaciens* cell.

In other embodiments, the host cell is a *Trichoderma koningii* cell, a *Trichoderma viride* cell, a *Trichoderma reesei* cell, a *Trichoderma longibrachiatum* cell, an *Aspergillus awamori* cell, an *Aspergillus fumigates* cell, an *Aspergillus foetidus* cell, an *Aspergillus nidulans* cell, an *Aspergillus niger* cell, an *Aspergillus oryzae* cell, a *Humicola insolens* cell, a *Humicola lanuginose* cell, a *Rhodococcus opacus* cell, a *Rhizomucor miehei* cell, or a *Mucor michei* cell.

In yet other embodiments, the host cell is a *Streptomyces lividans* cell or a *Streptomyces murinus* cell.

In yet other embodiments, the host cell is an *Actinomycetes* cell.

In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell. In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell.

In particular embodiments, the host cell is a cell from an eukaryotic plant, algae, cyanolacterium, green-sulfur bacterium, green non-sulfur bacterium, purple sulfur bacterium, purple non-sulfur bacterium, extremophile, yeast, fungus, engineered organisms thereof, or a synthetic organism. In some embodiments, the host cell is light dependent or fixes carbon. In some embodiments, the host cell is light dependent or fixes carbon. In some embodiments, the host cell has autotrophic activity. In some embodiments, the host cell has photoautotrophic activity, such as in the presence of light. In some embodiments, the host cell is heterotrophic or mixotrophic in the absence of light. In certain embodiments, the host cell is a cell from *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, *Zea mays*, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Synechococcus* Sp. PCC 7002, *Synechococcus* Sp. PCC 7942, *Synechocystis* Sp. PCC 6803, *Thermosynechococcus elongatus* BP-1, *Chlorobium tepidum*, *Chloroflexus auranticus*, *Chromatium vinosum*, *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Rhodopseudomonas palustris*, *Clostridium ljungdahlii*, *Clostridiothermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens*, or *Zymomonas mobilis*.

In other embodiments, the host cell is a CHO cell, a COS cell, a VERO cell, a BHK cell, a HeLa cell, a Cv1 cell, an MDCK cell, a 293 cell, a 3T3 cell, or a PC12 cell.

In yet other embodiments, the host cell is an *E. coli* cell. In certain embodiments, the *E. coli* cell is a strain B, a strain C, a strain K, or a strain W *E. coli* cell.

In another aspect, the invention features a method of producing a fatty alcohol. The method includes contacting a substrate with (i) a fatty alcohol biosynthetic polypeptide comprising the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194, or a variant thereof, or (ii) a fatty alcohol biosynthetic polypeptide encoded by a nucleotide sequence having at least about 70% identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193, or a variant thereof. In some embodiments, the method further includes purifying the fatty alcohol.

In some embodiments, the fatty alcohol biosynthetic polypeptide comprises the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194 with one or more amino acid substitutions, additions, insertions, or deletions, wherein the polypeptide has alcohol dehydrogenase activity.

In some embodiments, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has alcohol dehydrogenase activity.

In some embodiments, the polypeptide has an amino acid sequence that is at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194. In some embodiments, the polypeptide has the amino acid sequence of SEQ ID NO:94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, or 194.

In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193. In some embodiments, the nucleotide sequence is SEQ ID NO:93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, or 193.

In any of the aspects of the invention described herein, the methods can produce fatty alcohols comprising a C<sub>6</sub>-C<sub>26</sub> fatty alcohol. In some embodiments, the fatty alcohol comprises a C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>, C<sub>11</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, C<sub>18</sub>, C<sub>19</sub>, C<sub>20</sub>, C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>, C<sub>24</sub>, C<sub>25</sub>, or a C<sub>26</sub> fatty alcohol. In particular embodiments, the fatty alcohol is a C<sub>6</sub>, C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, or C<sub>18</sub> fatty alcohol. In certain embodiments, the hydroxyl group of the fatty alcohol is in the primary (C<sub>1</sub>) position.

In other embodiments, the fatty alcohol comprises a straight chain fatty alcohol. In other embodiments, the fatty alcohol comprises a branched chain fatty alcohol. In yet other embodiments, the fatty alcohol comprises a cyclic moiety.

In some embodiments, the fatty alcohol is an unsaturated fatty alcohol. In other embodiments, the fatty alcohol is a monounsaturated fatty alcohol. In certain embodiments, the unsaturated fatty alcohol is a C6:1, C7:1, C8:1, C9:1, C10:1, C11:1, C12:1, C13:1, C14:1, C15:1, C16:1, C17:1, C18:1, C19:1, C20:1, C21:1, C22:1, C23:1, C24:1, C25:1, or a C26:1 unsaturated fatty alcohol. In yet other embodiments, the fatty alcohol is unsaturated at the omega-7 position. In certain embodiments, the unsaturated fatty alcohol comprises a cis double bond.

In yet other embodiments, the fatty alcohol is a saturated fatty alcohol.

In any of the aspects of the invention described herein, a substrate for a fatty aldehyde biosynthetic polypeptide can be a fatty acid. In some embodiments, the fatty acid comprises a C<sub>6</sub>-C<sub>26</sub> fatty acid. In some embodiments, the fatty acid comprises a C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>, C<sub>11</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, C<sub>18</sub>, C<sub>19</sub>, C<sub>20</sub>, C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>, C<sub>24</sub>, C<sub>25</sub>, or a C<sub>26</sub> fatty acid. In particular embodiments, the fatty acid is a C<sub>6</sub>, C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, or C<sub>18</sub> fatty acid.

In other embodiments, the fatty acid comprises a straight chain fatty acid. In other embodiments, the fatty acid comprises a branched chain fatty acid. In yet other embodiments, the fatty acid comprises a cyclic moiety.

In some embodiments, the fatty acid is an unsaturated fatty acid. In other embodiments, the fatty acid is a monounsaturated fatty acid. In yet other embodiments, the fatty acid is a saturated fatty acid.

In another aspect, the invention features a genetically engineered microorganism comprising an exogenous control sequence stably incorporated into the genomic DNA of the microorganism upstream of a polynucleotide comprising a nucleotide sequence having at least about 70% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141,

143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271, wherein the microorganism produces an increased level of a fatty alcohol relative to a wild-type microorganism.

In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to the nucleotide sequence of SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271. In some embodiments, the nucleotide sequence is SEQ ID NO:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 263, 265, 267, 269, or 271.

In some embodiments, the polynucleotide is endogenous to the microorganism.

In other embodiments, the microorganism is genetically engineered to express a modified level of a gene encoding a fatty acid synthase in the host cell. In certain embodiments, the microorganism expresses a recombinant gene encoding a fatty acid synthase or expresses an increased level of an endogenous fatty acid synthase. In alternate embodiments, the microorganism expresses an attenuated level of a gene encoding a fatty acid synthase in the host cell and/or a decreased expression or activity of an endogenous fatty acid synthase. In some embodiments, the fatty acid synthase is a thioesterase. In particular embodiments, the thioesterase is encoded by tesA, tesA without leader sequence, tesB, fatB, fatB2, fatB3, fatA, or fatA1.

In other embodiments, the microorganism is genetically engineered to express an attenuated level of a fatty acid degradation enzyme relative to a wild type microorganism. In some embodiments, the microorganism expresses an attenuated level of an acyl-CoA synthase relative to a wild type microorganism. In particular embodiments, the microorganism expresses an attenuated level of an acyl-CoA synthase encoded by fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. In certain embodiments, the microorganism comprises a knockout of one or more genes encoding a fatty acid degradation enzyme, such as the aforementioned acyl-CoA synthase genes.

In yet other embodiments, the microorganism is genetically engineered to express an attenuated level of a dehydratase/isomerase enzyme, such as an enzyme encoded by fabA or by a gene listed in FIG. 15. In some embodiments, the microorganism comprises a knockout of fabA or a gene listed in FIG. 15. In other embodiments, the microorganism is genetically engineered to express an attenuated level of a ketoacyl-ACP synthase, such as an enzyme encoded by fabB or by a gene listed in FIG. 16. In certain embodiments, the microorganism comprises a knockout of fabB or a gene listed in FIG. 16. In yet other embodiments, the microorganism is

genetically engineered to express a modified level of a gene encoding a desaturase enzyme, such as desA.

In some embodiments, the microorganism is a bacterium. In certain embodiments, the bacterium is a Gram-negative or a Gram-positive bacterium.

In some embodiments, the microorganism is a mycobacterium selected from the group consisting of *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum*, and *Mycobacterium ulcerans*.

In other embodiments, the microorganism is *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinospira arenicola*, or *Clavibacter michiganensis*.

In another aspect, the invention features a fatty alcohol produced by any of the methods or any of the microorganisms described herein, or a surfactant comprising a fatty alcohol produced by any of the methods or any of the microorganisms described herein.

In some embodiments, the fatty alcohol has a  $\delta^{13}\text{C}$  of about -15.4 or greater. In certain embodiments, the fatty alcohol has a  $\delta^{13}\text{C}$  of about -15.4 to about -10.9, or of about -13.92 to about -13.84.

In some embodiments, the fatty alcohol has an  $f_M^{14}\text{C}$  of at least about 1.003. In certain embodiments, the fatty alcohol has an  $f_M^{14}\text{C}$  of at least about 1.01 or at least about 1.5. In some embodiments, the fatty alcohol has an  $f_M^{14}\text{C}$  of about 1.111 to about 1.124.

In any of the aspects described herein, a fatty alcohol is produced at a yield of about 25 mg/L, about 50 mg/L, about 75 mg/L, about 100 mg/L, about 125 mg/L, about 150 mg/L, about 175 mg/L, about 200 mg/L, about 225 mg/L, about 250 mg/L, about 275 mg/L, about 300 mg/L, about 325 mg/L, about 350 mg/L, about 375 mg/L, about 400 mg/L, about 425 mg/L, about 450 mg/L, about 475 mg/L, about 500 mg/L, about 525 mg/L, about 550 mg/L, about 575 mg/L, about 600 mg/L, about 625 mg/L, about 650 mg/L, about 675 mg/L, about 700 mg/L, about 725 mg/L, about 750 mg/L, about 775 mg/L, about 800 mg/L, about 825 mg/L, about 850 mg/L, about 875 mg/L, about 900 mg/L, about 925 mg/L, about 950 mg/L, about 975 mg/L, about 1000 g/L, about 1050 mg/L, about 1075 mg/L, about 1100 mg/L, about 1125 mg/L, about 1150 mg/L, about 1175 mg/L, about 1200 mg/L, about 1225 mg/L, about 1250 mg/L, about 1275 mg/L, about 1300 mg/L, about 1325 mg/L, about 1350 mg/L, about 1375 mg/L, about 1400 mg/L, about 1425 mg/L, about 1450 mg/L, about 1475 mg/L, about 1500 mg/L, about 1525 mg/L, about 1550 mg/L, about 1575 mg/L, about 1600 mg/L, about 1625 mg/L, about 1650 mg/L, about 1675 mg/L, about 1700 mg/L, about 1725 mg/L, about 1750 mg/L, about 1775 mg/L, about 1800 mg/L, about 1825 mg/L, about 1850 mg/L, about 1875 mg/L, about 1900 mg/L, about 1925 mg/L, about 1950 mg/L, about 1975 mg/L, about 2000 mg/L, or more.

In another aspect, the invention features a method of making a fatty alcohol described herein. The method includes culturing a host cell described herein in a medium having a low level of iron, under conditions sufficient to produce a fatty alcohol, as described herein. In particular embodiments, the medium contains less than about 500  $\mu\text{M}$  iron, less than about 400  $\mu\text{M}$  iron, less than about 300  $\mu\text{M}$  iron, less than about 200  $\mu\text{M}$  iron, less than about 150  $\mu\text{M}$  iron, less than about 100  $\mu\text{M}$  iron, less than about 90  $\mu\text{M}$  iron, less than about 80  $\mu\text{M}$  iron, less than about 70  $\mu\text{M}$  iron, less than about 60  $\mu\text{M}$  iron, less than about 50  $\mu\text{M}$  iron, less than about 40  $\mu\text{M}$  iron, less than about 30  $\mu\text{M}$  iron, less than about 20  $\mu\text{M}$  iron, less than about 10  $\mu\text{M}$  iron, or less than about 5  $\mu\text{M}$  iron. In certain embodiments, the medium does not contain iron.

In any of the aspects described herein, a fatty alcohol is produced in a host cell or a microorganism described herein from a carbon source.

The following figures are presented for the purpose of illustration only, and are not intended to be limiting.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a graphic representation of fatty alcohols produced by recombinant *E. coli* strains transformed with various plasmids.

FIG. 2 is a graphic representation of two GC/MS traces of organic compounds produced by recombinant *E. coli* strains transformed with various plasmids.

FIG. 3 is a schematic of a new pathway for fatty alcohol production.

FIG. 4 is a representation of a gel of PCR products from MG1655 wild type cells,  $\Delta fadD::cm$  cells, and  $\Delta fadD$  cells.

FIG. 5A is a GC/MS trace of fatty alcohol production in MG1655(DE3,  $\Delta fadD$ )/pETDUet-1-tesA+pHZ1.140B cells.

FIG. 5B is a GC/MS trace of fatty alcohol production in MG16655(DE3,  $\Delta fadD$ ,  $\gamma jgB::kan$ )/pETDUet-1-tesA+pHZ1.140B cells. FIG. 5C is a GC/MS trace of fatty alcohol production in MG16655(DE3,  $\Delta fadD$ ,  $\gamma jgB::kan$ )/pDF1+pHZ1.140B cells. The arrows in FIG. 5A, FIG. 5B, and FIG. 5C indicate the absence of C12:0 fatty aldehydes.

FIG. 6A-FIG. 6B is a listing of the nucleotide sequence and the corresponding amino acid sequence of *Nocardia* sp. NRRL 5646 car gene.

FIG. 7A-FIG. 7B is a listing of amino acid sequence motifs for CAR homologs.

FIG. 8A-FIG. 8UUU is a listing of nucleotide and amino acid sequences of car homolog genes.

FIG. 9 is a table identifying exemplary genes that can be expressed, overexpressed, or attenuated to increase production of particular substrates.

FIG. 10 is a listing of nucleotide and amino acid sequences of alcohol dehydrogenase genes.

FIG. 11 is a graphic representation of fatty alcohol production in various deletion mutants of *E. coli*.

FIG. 12 is a graphic representation of fatty alcohol production in various deletion mutants of *E. coli*.

FIG. 13 is a GC/MS trace of saturated fatty alcohol production in *E. coli*.

FIG. 14A is a graphic representation of fatty alcohol production in various Hu9 culture media. FIG. 14B is a graphic representation of fatty alcohol production in various Hu9 culture media.

FIG. 15 is a listing of nucleotide and amino acid sequences of fabA related genes.

FIG. 16 is a listing of nucleotide and amino acid sequences of fabB related genes.

FIG. 17 is a listing of additional nucleotide and amino acid sequences of the disclosure.

#### DETAILED DESCRIPTION OF THE INVENTION

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein, including GenBank database sequences, are incorporated by reference in their entirety. In case of conflict, the present

29

specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Definitions

Throughout the specification, a reference may be made using an abbreviated gene name or polypeptide name, but it is understood that such an abbreviated gene or polypeptide name represents the genus of genes or polypeptides. Such gene names include all genes encoding the same polypeptide and homologous polypeptides having the same physiological function. Polypeptide names include all polypeptides that have the same activity (e.g., that catalyze the same fundamental chemical reaction).

Unless otherwise indicated, the accession numbers referenced herein are derived from the NCBI database (National Center for Biotechnology Information) maintained by the National Institute of Health, U.S.A. Unless otherwise indicated, the accession numbers are as provided in the database as of October 2008.

EC numbers are established by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (available at on the world-wide web at [www.chem.qmul.ac.uk/iubmb/enzyme/](http://www.chem.qmul.ac.uk/iubmb/enzyme/)). The EC numbers referenced herein are derived from the KEGG Ligand database, maintained by the Kyoto Encyclopedia of Genes and Genomics, sponsored in part by the University of Tokyo. Unless otherwise indicated, the EC numbers are as provided in the database as of October 2008.

The articles “a” and “an” are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.

The term “about” is used herein to mean a value  $\pm 20\%$  of a given numerical value. Thus, “about 60%” means a value of between  $60 \pm (20\% \text{ of } 60)$  (i.e., between 48 and 70).

As used herein, the term “alcohol dehydrogenase” (EC 1.1.1.\*<sup>1</sup>) is a peptide capable of catalyzing the conversion of a fatty aldehyde to an alcohol (e.g., fatty alcohol). Additionally, one of ordinary skill in the art will appreciate that some alcohol dehydrogenases will catalyze other reactions as well. For example, some alcohol dehydrogenases will accept other substrates in addition to fatty aldehydes. Such non-specific alcohol dehydrogenases are, therefore, also included in this definition. Nucleic acid sequences encoding alcohol dehydrogenases are known in the art, and such alcohol dehydrogenases are publicly available. Exemplary GenBank Accession Numbers are provided in FIG. 9.

As used herein, the term “attenuate” means to weaken, reduce or diminish. For example, a polypeptide can be attenuated by modifying the polypeptide to reduce its activity (e.g., by modifying a nucleotide sequence that encodes the polypeptide).

As used herein, the term “biodiesel” means a biofuel that can be a substitute of diesel, which is derived from petroleum. Biodiesel can be used in internal combustion diesel engines in either a pure form, which is referred to as “neat” biodiesel, or as a mixture in any concentration with petroleum-based diesel. Biodiesel can include esters or hydrocarbons, such as alcohols.

As used therein, the term “biofuel” refers to any fuel derived from biomass. Biofuels can be substituted for petroleum based fuels. For example, biofuels are inclusive of trans-

30

portation fuels (e.g., gasoline, diesel, jet fuel, etc.), heating fuels, and electricity-generating fuels. Biofuels are a renewable energy source.

As used herein, the term “biomass” refers to any biological material from which a carbon source is derived. In some instances, a biomass is processed into a carbon source, which is suitable for bioconversion. In other instances, the biomass may not require further processing into a carbon source. The carbon source can be converted into a biofuel. One exemplary source of biomass is plant matter or vegetation. For example, corn, sugar cane, or switchgrass can be used as biomass. Another non-limiting example of biomass is metabolic wastes, such as animal matter, for example cow manure. In addition, biomass may include algae and other marine plants. Biomass also includes waste products from industry, agriculture, forestry, and households. Examples of such waste products that can be used as biomass are fermentation waste, ensilage, straw, lumber, sewage, garbage, cellulosic urban waste, and food leftovers. Biomass also includes sources of carbon, such as carbohydrates (e.g., monosaccharides, disaccharides, or polysaccharides).

As used herein, the phrase “carbon source” refers to a substrate or compound suitable to be used as a source of carbon for prokaryotic or simple eukaryotic cell growth. Carbon sources can be in various forms, including, but not limited to polymers, carbohydrates, acids, alcohols, aldehydes, ketones, amino acids, peptides, and gases (e.g., CO and CO<sub>2</sub>). These include, for example, various monosaccharides, such as glucose, fructose, mannose, and galactose; oligosaccharides, such as fructo-oligosaccharide and galacto-oligosaccharide; polysaccharides such as xylose and arabinose; disaccharides, such as sucrose, maltose, and turanose; cellulosic material, such as methyl cellulose and sodium carboxymethyl cellulose; saturated or unsaturated fatty acid esters, such as succinate, lactate, and acetate; alcohols, such as ethanol, methanol, and glycerol, or mixtures thereof. The carbon source can also be a product of photosynthesis, including, but not limited to, glucose. A preferred carbon source is biomass. Another preferred carbon source is glucose.

40 A nucleotide sequence is “complementary” to another nucleotide sequence if each of the bases of the two sequences matches (i.e., is capable of forming Watson Crick base pairs). The term “complementary strand” is used herein interchangeably with the term “complement”. The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand.

As used herein, a “cloud point lowering additive” is an additive added to a composition to decrease or lower the cloud point of a solution.

50 As used herein, the phrase “cloud point of a fluid” means the temperature at which dissolved solids are no longer completely soluble. Below this temperature, solids begin precipitating as a second phase giving the fluid a cloudy appearance. In the petroleum industry, cloud point refers to the temperature below which a solidified material or other heavy hydrocarbon crystallizes in a crude oil, refined oil, or fuel to form a cloudy appearance. The presence of solidified materials influences the flowing behavior of the fluid, the tendency of the fluid to clog fuel filters, injectors, etc., the accumulation of solidified materials on cold surfaces (e.g., a pipeline or heat exchanger fouling), and the emulsion characteristics of the fluid with water.

55 As used herein, the term “conditions sufficient to allow expression” means any conditions that allow a host cell to produce a desired product, such as a polypeptide or fatty aldehyde described herein. Suitable conditions include, for example, fermentation conditions. Fermentation conditions

31

can comprise many parameters, such as temperature ranges, levels of aeration, and media composition. Each of these conditions, individually and in combination, allows the host cell to grow. Exemplary culture media include broths or gels. Generally, the medium includes a carbon source, such as glucose, fructose, cellulose, or the like, that can be metabolized by a host cell directly. In addition, enzymes can be used in the medium to facilitate the mobilization (e.g., the depolymerization of starch or cellulose to fermentable sugars) and subsequent metabolism of the carbon source.

To determine if conditions are sufficient to allow expression, a host cell can be cultured, for example, for about 4, 8, 12, 24, 36, or 48 hours. During and/or after culturing, samples can be obtained and analyzed to determine if the conditions allow expression. For example, the host cells in the sample or the medium in which the host cells were grown can be tested for the presence of a desired product. When testing for the presence of a product, assays, such as, but not limited to, Thin Layer Chromatography (TLC), High-Performance Liquid Chromatography (HPLC), GC with Flame Ionization Detector (GC/FID), Gas chromatography—mass spectrometry (GC/MS), Liquid chromatography—mass spectrometry (LC/MS), mass spectrometry (MS), can be used.

It is understood that the polypeptides described herein may have additional conservative or non-essential amino acid substitutions, which do not have a substantial effect on the polypeptide functions. Whether or not a particular substitution will be tolerated (i.e., will not adversely affect desired biological properties, such as carboxylic acid reductase activity) can be determined as described in Bowie et al. *Science* (1990) 247:1306 1310. A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

As used herein, “control element” means a transcriptional control element. Control elements include promoters and enhancers. The term “promoter element,” “promoter,” or “promoter sequence” refers to a DNA sequence that functions as a switch that activates the expression of a gene. If the gene is activated, it is said to be transcribed or participating in transcription. Transcription involves the synthesis of mRNA from the gene. A promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA. Control elements interact specifically with cellular proteins involved in transcription (Maniatis et al., *Science* 236:1237, 1987).

As used herein, the term “fatty acid” means a carboxylic acid having the formula RCOOH. R represents an aliphatic group, preferably an alkyl group. R can comprise between about 4 and about 22 carbon atoms. Fatty acids can be saturated, monounsaturated, or polyunsaturated. In a preferred embodiment, the fatty acid is made from a fatty acid biosynthetic pathway.

As used herein, the term “fatty acid biosynthetic pathway” means a biosynthetic pathway that produces fatty acids. The fatty acid biosynthetic pathway includes fatty acid synthases that can be engineered, as described herein, to produce fatty

32

acids, and in some embodiments can be expressed with additional enzymes to produce fatty acids having desired carbon chain characteristics.

As used herein, the term “fatty acid degradation enzyme” means an enzyme involved in the breakdown or conversion of a fatty acid or fatty acid derivative into another product. A nonlimiting example of a fatty acid degradation enzyme is an acyl-CoA synthase. Additional examples of fatty acid degradation enzymes are described herein.

As used herein, the term “fatty acid derivative” means products made in part from the fatty acid biosynthetic pathway of the production host organism. “Fatty acid derivative” also includes products made in part from acyl-ACP or acyl-ACP derivatives. The fatty acid biosynthetic pathway includes fatty acid synthase enzymes which can be engineered as described herein to produce fatty acid derivatives, and in some examples can be expressed with additional enzymes to produce fatty acid derivatives having desired carbon chain characteristics. Exemplary fatty acid derivatives include for example, fatty acids, acyl-CoA, fatty aldehyde, short and long chain alcohols, hydrocarbons, fatty alcohols, and esters (e.g., waxes, fatty acid esters, or fatty esters).

As used herein, the term “fatty acid derivative enzyme” means any enzyme that may be expressed or overexpressed in the production of fatty acid derivatives. These enzymes may be part of the fatty acid biosynthetic pathway. Non-limiting examples of fatty acid derivative enzymes include fatty acid synthases, thioesterases, acyl-CoA synthases, acyl-CoA reductases, alcohol dehydrogenases, alcohol acyltransferases, fatty alcohol-forming acyl-CoA reductases, fatty acid (carboxylic acid) reductases, acyl-ACP reductases, fatty acid hydroxylases, acyl-CoA desaturases, acyl-ACP desaturases, acyl-CoA oxidases, acyl-CoA dehydrogenases, ester synthases, and alkane biosynthetic polypeptides, etc. Fatty acid derivative enzymes can convert a substrate into a fatty acid derivative. In some examples, the substrate may be a fatty acid derivative that the fatty acid derivative enzyme converts into a different fatty acid derivative.

As used herein, “fatty acid enzyme” means any enzyme involved in fatty acid biosynthesis. Fatty acid enzymes can be modified in host cells to produce fatty acids. Non-limiting examples of fatty acid enzymes include fatty acid synthases and thioesterases. Additional examples of fatty acid enzymes are described herein.

As used herein, “fatty acid synthase” means any enzyme involved in fatty acid biosynthesis. Fatty acid synthases can be expressed or overexpressed in host cells to produce fatty acids. A non-limiting example of a fatty acid synthase is a thioesterase. Additional examples of fatty acid synthases are described herein.

As used herein, “fatty aldehyde” means an aldehyde having the formula RCHO characterized by an unsaturated carbonyl group (C=O). In a preferred embodiment, the fatty aldehyde is any aldehyde made from a fatty acid or fatty acid derivative. In one embodiment, the R group is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbons in length.

R can be straight or branched chain. The branched chains may have one or more points of branching. In addition, the branched chains may include cyclic branches.

Furthermore, R can be saturated or unsaturated. If unsaturated, the R can have one or more points of unsaturation.

In one embodiment, the fatty aldehyde is produced biosynthetically.

Fatty aldehydes have many uses. For example, fatty aldehydes can be used to produce many specialty chemicals. For example, fatty aldehydes are used to produce polymers, res-

ins, dyes, flavorings, plasticizers, perfumes, pharmaceuticals, and other chemicals. Some are used as solvents, preservatives, or disinfectants. Some natural and synthetic compounds, such as vitamins and hormones, are aldehydes.

The terms "fatty aldehyde biosynthetic polypeptide", "carboxylic acid reductase", and "CAR" are used interchangeably herein.

As used herein, "fatty alcohol" means an alcohol having the formula ROH. In a preferred embodiment, the fatty alcohol is any alcohol made from a fatty acid or fatty acid derivative. In one embodiment, the R group is at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbons in length.

R can be straight or branched chain. The branched chains may have one or more points of branching. In addition, the branched chains may include cyclic branches.

Furthermore, R can be saturated or unsaturated. If unsaturated, the R can have one or more points of unsaturation.

In one embodiment, the fatty alcohol is produced biosynthetically.

Fatty alcohols have many uses. For example, fatty alcohols can be used to produce many specialty chemicals. For example, fatty alcohols are used as a biofuel; as solvents for fats, waxes, gums, and resins; in pharmaceutical salves, emollients and lotions; as lubricating-oil additives; in detergents and emulsifiers; as textile antistatic and finishing agents; as plasticizers; as nonionic surfactants; and in cosmetics, for examples as thickeners.

As used herein, "fraction of modern carbon" or " $f_M$ " has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOxI and HOxII, respectively. The fundamental definition relates to 0.95 times the  $^{14}\text{C}/^{12}\text{C}$  isotope ratio HOxI (referenced to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (plant material),  $f_M$  is approximately 1.1.

"Gene knockout", as used herein, refers to a procedure by which a gene encoding a target protein is modified or inactivated so to reduce or eliminate the function of the intact protein. Inactivation of the gene may be performed by general methods such as mutagenesis by UV irradiation or treatment with N-methyl-N'-nitro-N-nitrosoguanidine, site-directed mutagenesis, homologous recombination, insertion-deletion mutagenesis, or "Red-driven integration" (Datsenko et al., *Proc. Natl. Acad. Sci. USA*, 97:6640-45, 2000). For example, in one embodiment, a construct is introduced into a host cell, such that it is possible to select for homologous recombination events in the host cell. One of skill in the art can readily design a knock-out construct including both positive and negative selection genes for efficiently selecting transfected cells that undergo a homologous recombination event with the construct. The alteration in the host cell may be obtained, for example, by replacing through a single or double cross-over recombination a wild type DNA sequence by a DNA sequence containing the alteration. For convenient selection of transformants, the alteration may, for example, be a DNA sequence encoding an antibiotic resistance marker or a gene complementing a possible auxotrophy of the host cell. Mutations include, but are not limited to, deletion-insertion mutations. An example of such an alteration includes a gene disruption, i.e., a perturbation of a gene such that the product that is normally produced from this gene is not produced in a functional form. This could be due to a complete deletion, a deletion and insertion of a selective marker, an insertion of a selective marker, a frameshift mutation, an in-frame deletion, or a point mutation that leads to premature termination. In

some instances, the entire mRNA for the gene is absent. In other situations, the amount of mRNA produced varies.

Calculations of "homology" between two sequences can be performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence that is aligned for comparison purposes is at least about 30%, preferably at least about 40%, more preferably at least about 50%, even more preferably at least about 60%, and even more preferably at least about 70%, at least about 80%, at least about 90%, or about 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent homology between two amino acid sequences is determined using the Needleman and Wunsch (1970), *J. Mol. Biol.* 48:444 453, algorithm that has been incorporated into the GAP program in the GCG software package, using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent homology between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdn.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about which parameters should be applied to determine if a molecule is within a homology limitation of the claims) are a Blosum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

As used herein, a "host cell" is a cell used to produce a product described herein (e.g., a fatty alcohol described herein). A host cell can be modified to express or overexpress selected genes or to have attenuated expression of selected genes. Non-limiting examples of host cells include plant, animal, human, bacteria, yeast, or filamentous fungi cells.

As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either method can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by two washes in 0.2×SSC, 0.1% SDS at least at 50° C. (the temperature of the washes can be increased to 55° C. for low stringency conditions); 2) medium stringency hybridization conditions in 6×SSC at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 60° C.; 3) high

stringency hybridization conditions in 6×SSC at about 45°C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 65° C.; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65° C., followed by one or more washes at 0.2×SSC, 1% SDS at 65° C. Very high stringency conditions (4) are the preferred conditions unless otherwise specified.

The term “isolated” as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the nucleic acid. Moreover, by an “isolated nucleic acid” is meant to include nucleic acid fragments, which are not naturally occurring as fragments and would not be found in the natural state. The term “isolated” is also used herein to refer to polypeptides, which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. The term “isolated” as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques. The term “isolated” as used herein also refers to a nucleic acid or peptide that is substantially free of chemical precursors or other chemicals when chemically synthesized. The term “isolated”, as used herein with respect to products, such as fatty alcohols, refers to products that are isolated from cellular components, cell culture media, or chemical or synthetic precursors.

As used herein, the “level of expression of a gene in a cell” refers to the level of mRNA, pre-mRNA nascent transcript(s), transcript processing intermediates, mature mRNA(s), and degradation products encoded by the gene in the cell.

As used herein, the term “microorganism” means prokaryotic and eukaryotic microbial species from the domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The terms “microbial cells” (i.e., cells from microbes) and “microbes” are used interchangeably and refer to cells or small organisms that can only be seen with the aid of a microscope.

As used herein, the term “nucleic acid” refers to polynucleotides, such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides, ESTs, chromosomes, cDNAs, mRNAs, and rRNAs.

As used herein, the term “operably linked” means that selected nucleotide sequence (e.g., encoding a polypeptide described herein) is in proximity with a promoter to allow the promoter to regulate expression of the selected DNA. In addition, the promoter is located upstream of the selected nucleotide sequence in terms of the direction of transcription and translation. By “operably linked” is meant that a nucleotide sequence and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

The term “or” is used herein to mean, and is used interchangeably with, the term “and/or,” unless context clearly indicates otherwise.

As used herein, “overexpress” means to express or cause to be expressed a nucleic acid, polypeptide, or hydrocarbon in a cell at a greater concentration than is normally expressed in a corresponding wild-type cell. For example, a polypeptide can be “overexpressed” in a recombinant host cell when the polypeptide is present in a greater concentration in the recom-

binant host cell compared to its concentration in a non-recombinant host cell of the same species.

As used herein, “partition coefficient” or “P,” is defined as the equilibrium concentration of a compound in an organic phase divided by the concentration at equilibrium in an aqueous phase (e.g., fermentation broth). In one embodiment of a bi-phasic system described herein, the organic phase is formed by the fatty aldehyde during the production process. However, in some examples, an organic phase can be provided, such as by providing a layer of octane, to facilitate product separation. When describing a two phase system, the partition characteristics of a compound can be described as log P. For example, a compound with a log P of 1 would partition 10:1 to the organic phase. A compound with a log P of -1 would partition 1:10 to the organic phase. By choosing an appropriate fermentation broth and organic phase, a fatty aldehyde with a high log P value can separate into the organic phase even at very low concentrations in the fermentation vessel.

As used herein, the term “purify,” “purified,” or “purification” means the removal or isolation of a molecule from its environment by, for example, isolation or separation. “Substantially purified” molecules are at least about 60% free, preferably at least about 75% free, and more preferably at least about 90% free from other components with which they are associated. As used herein, these terms also refer to the removal of contaminants from a sample. For example, the removal of contaminants can result in an increase in the percentage of fatty alcohol in a sample. For example, when fatty alcohols are produced in a host cell, the fatty alcohols can be purified by the removal of host cell proteins. After purification, the percentage of fatty alcohols in the sample is increased.

The terms “purify,” “purified,” and “purification” do not require absolute purity. They are relative terms. Thus, for example, when fatty alcohols are produced in host cells, a purified fatty alcohol is one that is substantially separated from other cellular components (e.g., nucleic acids, polypeptides, lipids, carbohydrates, or other hydrocarbons). In another example, a purified fatty alcohol preparation is one in which the fatty alcohol is substantially free from contaminants, such as those that might be present following fermentation. In some embodiments, a fatty alcohol is purified when at least about 50% by weight of a sample is composed of the fatty alcohol. In other embodiments, a fatty alcohol is purified when at least about 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% or more by weight of a sample is composed of the fatty alcohol.

As used herein, the term “recombinant polypeptide” refers to a polypeptide that is produced by recombinant DNA techniques, wherein generally DNA encoding the expressed protein or RNA is inserted into a suitable expression vector and that is in turn used to transform a host cell to produce the polypeptide or RNA.

As used herein, the term “substantially identical” (or “substantially homologous”) is used to refer to a first amino acid or nucleotide sequence that contains a sufficient number of identical or equivalent (e.g., with a similar side chain, e.g., conserved amino acid substitutions) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have similar activities.

As used herein, the term “synthase” means an enzyme which catalyzes a synthesis process. As used herein, the term synthase includes synthases, synthetases, and ligases.

As used herein, the term "transfection" means the introduction of a nucleic acid (e.g., via an expression vector) into a recipient cell by nucleic acid-mediated gene transfer.

As used herein, "transformation" refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA. This may result in the transformed cell expressing a recombinant form of an RNA or polypeptide. In the case of antisense expression from the transferred gene, the expression of a naturally-occurring form of the polypeptide is disrupted.

As used herein, a "transport protein" is a polypeptide that facilitates the movement of one or more compounds in and/or out of a cellular organelle and/or a cell.

As used herein, a "variant" of polypeptide X refers to a polypeptide having the amino acid sequence of peptide X in which one or more amino acid residues is altered. The variant may have conservative changes or nonconservative changes. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without affecting biological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to that of a gene or the coding sequence thereof. This definition may also include, for example, "allelic," "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference polynucleotide, but will generally have a greater or fewer number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of useful vector is an episome (i.e., a nucleic acid capable of extra-chromosomal replication). Useful vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids," which refer generally to circular double stranded DNA loops that, in their vector form, are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably, as the plasmid is the most commonly used form of vector. However, also included are such other forms of expression vectors that serve equivalent functions and that become known in the art subsequently hereto.

The invention is based, at least in part, on the discovery of a new pathway for fatty alcohol biosynthesis in *E. coli* that utilize, in part, genes that encode fatty aldehyde biosynthetic polypeptides. The fatty alcohols can be produced by a biosynthetic pathway depicted in FIG. 3. In this pathway, a fatty acid is first activated by ATP and then reduced by a carboxylic acid reductase (CAR)-like enzyme to generate a fatty aldehyde. The fatty aldehyde can then be further reduced into a fatty alcohol by an alcohol dehydrogenase(s), such as *alrAadp1* or *yjgB*. As demonstrated herein, *yjgB* may be the presumed alcohol dehydrogenase, whose substrates includes

fatty aldehydes, for example fatty aldehydes with carbon chain lengths from C<sub>10</sub> to C<sub>18</sub>.

#### Fatty Aldehyde Biosynthetic Genes, Fatty Alcohol Biosynthetic Genes, and Variants

The methods described herein can be used to produce fatty alcohols, for example, from fatty aldehydes. In some instances, a fatty aldehyde is produced by expressing a fatty aldehyde biosynthetic gene, for example, a carboxylic acid reductase gene (car gene), having a nucleotide sequence listed in FIGS. 6 and 8, as well as polynucleotide variants thereof. In some instances, the fatty aldehyde biosynthetic gene encodes one or more of the amino acid motifs depicted in FIG. 7. For example, the gene can encode a polypeptide comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; and/or SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11. SEQ ID NO:7 includes a reductase domain; SEQ ID NO:8 and SEQ ID NO:14 include a NADP binding domain; SEQ ID NO:9 includes a phosphopantetheine attachment site; and SEQ ID NO:10 includes an AMP binding domain.

In other instances, a fatty alcohol is produced by expressing a fatty alcohol biosynthetic gene, for example, having a nucleotide sequence listed in FIG. 10, or a variant thereof.

Variants can be naturally occurring or created in vitro. In particular, such variants can be created using genetic engineering techniques, such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, or standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives can be created using chemical synthesis or modification procedures.

Methods of making variants are well known in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids that encode polypeptides having characteristics that enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Typically, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants can be created using error prone PCR (see, e.g., Leung et al., *Technique* 1:11-15, 1989; and Caldwell et al., *PCR Methods Appl.* 2:28-33, 1992). In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Briefly, in such procedures, nucleic acids to be mutagenized (e.g., a fatty aldehyde biosynthetic polynucleotide sequence), are mixed with PCR primers, reaction buffer, MgCl<sub>2</sub>, MnCl<sub>2</sub>, Taq polymerase, and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction can be performed using 20 fmoles of nucleic acid to be mutagenized (e.g., a fatty aldehyde biosynthetic polynucleotide sequence), 30 pmole of each PCR primer, a reaction buffer comprising 50 mM KCl, 10 mM Tris HCl (pH 8.3), and 0.01% gelatin, 7 mM MgCl<sub>2</sub>, 0.5 mM MnCl<sub>2</sub>, 5 units of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM dTTP. PCR can be performed for 30 cycles of 94° C. for 1 min, 45° C. for 1 min, and 72° C. for 1 min. However, it will be appreciated that these parameters can be varied as appropriate. The mutagenized nucleic acids are then cloned into an appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids are evaluated.

Variants can also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described in, for example, Reidhaar-Olson et al., *Science* 241:53-57, 1988. Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized (e.g., a fatty aldehyde biosynthetic polynucleotide sequence). Clones containing the mutagenized DNA are recovered, and the activities of the polypeptides they encode are assessed.

Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in, for example, U.S. Pat. No. 5,965,408.

Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different, but highly related, DNA sequence in vitro as a result of random fragmentation of the DNA molecule based on sequence homology. This is followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described in, for example, Stemmer, *PNAS, USA* 91:1074740751, 1994.

Variants can also be created by in vivo mutagenesis. In some embodiments, random mutations in a nucleic acid sequence are generated by propagating the sequence in a bacterial strain, such as an *E. coli* strain, which carries mutations in one or more of the DNA repair pathways. Such "mutator" strains have a higher random mutation rate than that of a wild-type strain. Propagating a DNA sequence (e.g., a fatty aldehyde biosynthetic polynucleotide sequence) in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for in vivo mutagenesis are described in, for example, PCT Publication No. WO 91/16427.

Variants can also be generated using cassette mutagenesis. In cassette mutagenesis, a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains a completely and/or partially randomized native sequence.

Recursive ensemble mutagenesis can also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (i.e., protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described in, for example, Arkin et al., *PNAS, USA* 89:7811-7815, 1992.

In some embodiments, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described in, for example, Delegrave et al., *Biotech. Res.* 11:1548-1552, 1993. Random and site-directed mutagenesis are described in, for example, Arnold, *Curr. Opin. Biotech.* 4:450-455, 1993.

In some embodiments, variants are created using shuffling procedures wherein portions of a plurality of nucleic acids

that encode distinct polypeptides are fused together to create chimeric nucleic acid sequences that encode chimeric polypeptides as described in, for example, U.S. Pat. Nos. 5,965,408 and 5,939,250.

Polynucleotide variants also include nucleic acid analogs. Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxyuridine for deoxymidine and 5-methyl-2'-deoxycytidine or 5-bromo-2'-deoxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six-membered, morpholino ring, or peptide nucleic acids, in which the deoxypyrophosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. (See, e.g., Summerton et al., *Antisense Nucleic Acid Drug Dev.* (1997) 7:187-195; and Hyrup et al., *Bioorgan. Med. Chem.* (1996) 4:5-23.) In addition, the deoxypyrophosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

Any polynucleotide sequence encoding a homolog listed in FIGS. 6 and 8, or a variant thereof, can be used as a fatty aldehyde biosynthetic polynucleotide in the methods described herein. Any polynucleotide sequence listed in FIG. 10, or a variant, can be used as a fatty alcohol biosynthetic polynucleotide in the methods described herein.

#### Fatty Aldehyde Biosynthetic Polypeptides, Fatty Alcohol Biosynthetic Polypeptide, and Variants

The methods described herein can also be used to produce fatty alcohols, for example, from fatty aldehydes. In some instances, the fatty aldehyde is produced by a fatty aldehyde biosynthetic polypeptide having an amino acid sequence listed in FIGS. 6 and 8, as well as polypeptide variants thereof. In some instances, a fatty aldehyde biosynthetic polypeptide is one that includes one or more of the amino acid motifs depicted in FIG. 7. For example, the polypeptide can include the amino acid sequences of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10. In other situations, the polypeptide includes one or more of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14. In yet other instances, the polypeptide includes the amino acid sequences of SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11. SEQ ID NO:7 includes a reductase domain; SEQ ID NO:8 and SEQ ID NO:14 include a NADP binding domain; SEQ ID NO:9 includes a phosphopantetheine attachment site; and SEQ ID NO:10 includes an AMP binding domain.

In other instances, the methods described herein can be used to produce fatty alcohols using a fatty alcohol biosynthetic polypeptide having an amino acid sequence listed in FIG. 10, as well as polypeptide variants thereof.

Biosynthetic polypeptide variants can be variants in which one or more amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue). Such substituted amino acid residue may or may not be one encoded by the genetic code.

Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of similar characteristics. Typical conservative substitutions are the following replacements: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine or vice versa; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue;

replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue.

Other polypeptide variants are those in which one or more amino acid residues include a substituent group. Still other polypeptide variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (e.g., polyethylene glycol).

Additional polypeptide variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretory sequence, a proprotein sequence, or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

In some instances, the polypeptide variants retain the same biological function as a polypeptide having an amino acid sequence listed in FIGS. 6 and 8 (e.g., retain fatty aldehyde biosynthetic activity, such as carboxylic acid or fatty acid reductase activity), or listed in FIG. 10 (e.g., retain fatty alcohol biosynthetic activity, such as fatty alcohol dehydrogenase activity) and have amino acid sequences substantially identical thereto.

In other instances, the polypeptide variants have at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or more than about 95% homology to an amino acid sequence listed in FIGS. 6, 8, and/or 10. In another embodiment, the polypeptide variants include a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof.

The polypeptide variants or fragments thereof can be obtained by isolating nucleic acids encoding them using techniques described herein or by expressing synthetic nucleic acids encoding them. Alternatively, polypeptide variants or fragments thereof can be obtained through biochemical enrichment or purification procedures. The sequence of polypeptide variants or fragments can be determined by proteolytic digestion, gel electrophoresis, and/or microsequencing. The sequence of the polypeptide variants or fragments can then be compared to an amino acid sequence listed in FIGS. 6, 8, and/or 10 using any of the programs described herein.

The polypeptide variants and fragments thereof can be assayed for fatty aldehyde-producing activity and/or fatty alcohol-producing activity using routine methods. For example, the polypeptide variants or fragment can be contacted with a substrate (e.g., a fatty acid, a fatty acid derivative substrate, or other substrate described herein) under conditions that allow the polypeptide variant to function. A decrease in the level of the substrate or an increase in the level of a fatty aldehyde can be measured to determine fatty aldehyde-producing activity. A decrease in the level of the substrate or an increase in the level of a fatty alcohol can be measured to determine fatty alcohol-producing activity.

#### Antibodies to Biosynthetic Polypeptides

The fatty aldehyde biosynthetic polypeptides described herein can also be used to produce antibodies directed against fatty aldehyde biosynthetic polypeptides. Such antibodies can be used, for example, to detect the expression of a fatty aldehyde biosynthetic polypeptide or fatty alcohol biosynthetic polypeptide using methods known in the art. The antibody can be, for example, a polyclonal antibody; a monoclonal antibody or antigen binding fragment thereof; a

modified antibody such as a chimeric antibody, reshaped antibody, humanized antibody, or fragment thereof (e.g., Fab', Fab, F(ab')<sub>2</sub>); or a biosynthetic antibody, for example, a single chain antibody, single domain antibody (DAB), Fv, single chain Fv (scFv), or the like.

Methods of making and using polyclonal and monoclonal antibodies are described, for example, in Harlow et al., *Using Antibodies: A Laboratory Manual: Portable Protocol I*. Cold Spring Harbor Laboratory (Dec. 1, 1998). Methods for making modified antibodies and antibody fragments (e.g., chimeric antibodies, reshaped antibodies, humanized antibodies, or fragments thereof, e.g., Fab', Fab, F(ab')<sub>2</sub> fragments); or biosynthetic antibodies (e.g., single chain antibodies, single domain antibodies (DABs), Fv, single chain Fv (scFv), and the like), are known in the art and can be found, for example, in Zola, *Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives*, Springer Verlag (Dec. 15, 2000; 1st edition). Substrates

The compositions and methods described herein can be used to produce fatty alcohols, for example, from fatty aldehydes, which themselves can be produced from an appropriate substrate. While not wishing to be bound by theory, it is believed that the fatty aldehyde biosynthetic polypeptides described herein produce fatty aldehydes from substrates via a reduction mechanism. In some instances, the substrate is a fatty acid derivative (e.g., a fatty acid), and a fatty aldehyde having particular branching patterns and carbon chain length can be produced from a fatty acid derivative having those characteristics that would result in a particular fatty aldehyde. Through an additional reaction mechanism, the fatty aldehyde can be converted into the desired fatty alcohol (e.g., by a fatty alcohol biosynthetic polypeptide described herein).

Accordingly, each step within a biosynthetic pathway that leads to the production of a fatty acid derivative substrate can be modified to produce or overproduce the substrate of interest. For example, known genes involved in the fatty acid biosynthetic pathway or the fatty aldehyde pathway can be expressed, overexpressed, or attenuated in host cells to produce a desired substrate (see, e.g., PCT/US08/058,788). Exemplary genes are provided in FIG. 9.

#### Synthesis of Substrates

Fatty acid synthase (FAS) is a group of polypeptides that catalyze the initiation and elongation of acyl chains (Marrackchi et al., *Biochemical Society*, 30:1050-1055, 2002). The acyl carrier protein (ACP) along with the enzymes in the FAS pathway control the length, degree of saturation, and branching of the fatty acid derivatives produced. The fatty acid biosynthetic pathway involves the precursors acetyl-CoA and malonyl-CoA. The steps in this pathway are catalyzed by enzymes of the fatty acid biosynthesis (fab) and acetyl-CoA carboxylase (acc) gene families (see, e.g., Heath et al., *Prog. Lipid Res.* 40(6):467-97 (2001)).

Host cells can be engineered to express fatty acid derivative substrates by recombinantly expressing or overexpressing one or more fatty acid synthase genes, such as acetyl-CoA and/or malonyl-CoA synthase genes. For example, to increase acetyl-CoA production, one or more of the following genes can be expressed in a host cell: pdh (a multienzyme complex comprising aceEF (which encodes the E1p dehydrogenase component, the E2p dihydrolipoamide acyltransferase component of the pyruvate and 2-oxoglutarate dehydrogenase complexes, and lpd), panK, fabH, fabB, fabD, fabG, acpP, and fabF. Exemplary GenBank accession numbers for these genes are: pdh (BAB34380, AAC73227, AAC73226), panK (also known as CoA, AAC76952), aceEF (AAC73227, AAC73226), fabH (AAC74175), fabB

(POA953), fabD (AAC74176), fabG (AAC74177), acpP (AAC74178), fabF (AAC74179). Additionally, the expression levels of fadE, gpsA, ldhA, pflb, adhE, pta, poxB, ackA, and/or ackB can be attenuated or knocked-out in an engineered host cell by transformation with conditionally replicative or non-replicative plasmids containing null or deletion mutations of the corresponding genes or by substituting promoter or enhancer sequences. Exemplary GenBank accession numbers for these genes are: fadE (AAC73325), gspA (AAC76632), ldhA (AAC74462), pflb (AAC73989), adhE (AAC74323), pta (AAC75357), poxB (AAC73958), ackA (AAC75356), and ackB (BAB81430). The resulting host cells will have increased acetyl-CoA production levels when grown in an appropriate environment.

Malonyl-CoA overexpression can be affected by introducing accABCD (e.g., accession number AAC73296, EC 6.4.1.2) into a host cell. Fatty acids can be further overexpressed in host cells by introducing into the host cell a DNA sequence encoding a lipase (e.g., accession numbers CAA89087, CAA98876).

In addition, inhibiting PlsB can lead to an increase in the levels of long chain acyl-ACP, which will inhibit early steps in the pathway (e.g., accABCD, fabH, and fabI). The plsB (e.g., accession number AAC77011) D311E mutation can be used to increase the amount of available fatty acids.

In addition, a host cell can be engineered to overexpress a sfa gene (suppressor of fabA, e.g., accession number AAN79592) to increase production of monounsaturated fatty acids (Rock et al., *J. Bacteriology* 178:5382-5387, 1996).

The chain length of a fatty acid derivative substrate can be selected for by modifying the expression of selected thioesterases. Thioesterase influences the chain length of fatty acids produced. Hence, host cells can be engineered to express, overexpress, have attenuated expression, or not to express one or more selected thioesterases to increase the production of a preferred fatty acid derivative substrate. For example, C<sub>10</sub> fatty acids can be produced by expressing a thioesterase that has a preference for producing C<sub>10</sub> fatty acids and attenuating thioesterases that have a preference for producing fatty acids other than C<sub>10</sub> fatty acids (e.g., a thioesterase which prefers to produce C<sub>14</sub> fatty acids). This would result in a relatively homogeneous population of fatty acids that have a carbon chain length of 10. In other instances, C<sub>14</sub> fatty acids can be produced by attenuating endogenous thioesterases that produce non-C<sub>14</sub> fatty acids and expressing the thioesterases that use C<sub>14</sub>-ACP. In some situations, C<sub>12</sub> fatty acids can be produced by expressing thioesterases that use C<sub>12</sub>-ACP and attenuating thioesterases that produce non-C<sub>12</sub> fatty acids. Acetyl-CoA, malonyl-CoA, and fatty acid overproduction can be verified using methods known in the art, for example, by using radioactive precursors, HPLC, or GC-MS subsequent to cell lysis. Non-limiting examples of thioesterases that can be used in the methods described herein are listed in Table 1.

TABLE 1

Thioesterases		
Accession Number	Source Organism	Gene
AAC73596	<i>E. coli</i>	tesA without leader sequence
AAC73555	<i>E. coli</i>	tesB
Q41635, AAA34215	<i>Umbellularia californica</i>	fadB
AAC49269	<i>Cuphea hookeriana</i>	fadB2
Q39513; AAC72881	<i>Cuphea hookeriana</i>	fadB3
Q39473, AAC49151	<i>Cinnamomum camphorum</i>	fadB

TABLE 1-continued

Thioesterases		
Accession Number	Source Organism	Gene
CAA85388	<i>Arabidopsis thaliana</i>	fadB [M141T]*
NP_189147; NP_193041	<i>Arabidopsis thaliana</i>	fadA
CAC39106	<i>Bradyrhizobium japonicum</i>	fadA
AAC72883	<i>Cuphea hookeriana</i>	fadA
AAL79361	<i>Helianthus annus</i>	fadA1

\*Mayer et al., *BMC Plant Biology* 7: 1-11, 2007

In other instances, a fatty aldehyde biosynthetic polypeptide, variant, or a fragment thereof, is expressed in a host cell that contains a naturally occurring mutation that results in an increased level of fatty acids in the host cell. In some instances, the host cell is genetically engineered to increase the level of fatty acids in the host cell relative to a corresponding wild-type host cell. For example, the host cell can be genetically engineered to express a reduced level of an acyl-CoA synthase relative to a corresponding wild-type host cell. In one embodiment, the level of expression of one or more genes (e.g., an acyl-CoA synthase gene) is reduced by genetically engineering a “knock out” host cell.

Any known acyl-CoA synthase gene can be reduced or knocked out in a host cell. Non-limiting examples of acyl-CoA synthase genes include fadD, fadK, BH3103, yhfL, Pfl-4354, EAV15023, fadD1, fadD2, RPC\_4074, fadDD35, fadDD22, faa3p or the gene encoding the protein ZP\_01644857. Specific examples of acyl-CoA synthase genes include fadDD35 from *M. tuberculosis* H37Rv [NP\_217021], fadDD22 from *M. tuberculosis* H37Rv [NP\_217464], fadD from *E. coli* [NP\_416319], fadK from *E. coli* [YP\_416216], fadD from *Acinetobacter* sp. ADP1 [YP\_045024], fadD from *Haemophilus influenzae* RdKW20 [NP\_438551], fadD from *Rhodopseudomonas palustris* Bis B18 [YP\_533919], BH3101 from *Bacillus halodurans* C-125 [NP\_243969], Pfl-4354 from *Pseudomonas fluorescens* Pfo-1 [YP\_350082], EAV15023 from *Comamonas testosterone* KF-1 [ZP\_01520072], yhfL from *B. subtilis* [NP\_388908], fadD1 from *P. aeruginosa* PAO1 [NP\_251989], fadD1 from *Ralstonia solanacearum* GM1 1000 [NP\_520978], fadD2 from *P. aeruginosa* PAO1 [NP\_251990], the gene encoding the protein ZP\_01644857 from *Stenotrophomonas maltophilia* R551-3, faa3p from *Saccharomyces cerevisiae* [NP\_012257], faa1p from *Saccharomyces cerevisiae* [NP\_014962], lcfa from *Bacillus subtilis* [CAA99571], or those described in Shockley et al., *Plant. Physiol.* 129:1710-1722, 2002; Caviglia et al., *J. Biol. Chem.* 279:1163-1169, 2004; Knoll et al., *J. Biol. Chem.* 269(23):16348-56, 1994; Johnson et al., *J. Biol. Chem.* 269: 18037-18046, 1994; and Black et al., *J. Biol. Chem.* 267: 25513-25520, 1992.

#### Formation of Branched Fatty Alcohols

Fatty alcohols can be produced from fatty aldehydes that contain branch points by using branched fatty acid derivatives as substrates for a fatty aldehyde biosynthetic polypeptide described herein. For example, although *E. coli* naturally produces straight chain fatty acids (SFAs), *E. coli* can be engineered to produce branched chain fatty acids (brFAs) by introducing and expressing or overexpressing genes that provide branched precursors in the *E. coli* (e.g., bkd, ilv, icm, and fab gene families). Additionally, a host cell can be engineered to express or overexpress genes encoding proteins for the elongation of brFAs (e.g., ACP, FabF, etc.) and/or to delete or attenuate the corresponding host cell genes that normally lead to SFAs.

The first step in forming brFAs is the production of the corresponding  $\alpha$ -keto acids by a branched-chain amino acid aminotransferase. Host cells may endogenously include genes encoding such enzymes or such genes can be recombinantly introduced. *E. coli*, for example, endogenously expresses such an enzyme, IlvE (EC 2.6.1.42; GenBank accession YP\_026247). In some host cells, a heterologous branched-chain amino acid aminotransferase may not be expressed. However, *E. coli* IlvE or any other branched-chain amino acid aminotransferase (e.g., IlvE from *Lactococcus lactis* (GenBank accession AAF34406), IlvE from *Pseudomonas putida* (GenBank accession NP\_745648), or IlvE from *Streptomyces coelicolor* (GenBank accession NP\_629657)), if not endogenous, can be introduced.

In another embodiment, the production of  $\alpha$ -keto acids can be achieved by using the methods described in Atsumi et al., *Nature* 451:86-89, 2008. For example, 2-ketoisovalerate can be produced by overexpressing the genes encoding IlvI, INH, IlvC, or IlvD. In another example, 2-keto-3-methyl-valerate can be produced by overexpressing the genes encoding IlvA and IlvI, IlvH (or AlsS of *Bacillus subtilis*), INC, IlvD, or their corresponding homologs. In a further embodiment, 2-keto-4-methyl-pentanoate can be produced by overexpressing the genes encoding IlvI, IlvH, IlvC, IlvD and LeuA, LeuB, LeuC, LeuD, or their corresponding homologs.

The second step is the oxidative decarboxylation of the  $\alpha$ -keto acids to the corresponding branched-chain acyl-CoA. This reaction can be catalyzed by a branched-chain  $\alpha$ -keto acid dehydrogenase complex (bkd; EC 1.2.4.4.) (Denoya et al., *J. Bacteriol.* 177:3504, 1995), which consists of E1 $\alpha/\beta$  (decarboxylase), E2 (dihydrolipoyl transacylase), and E3 (dihydrolipoyl dehydrogenase) subunits. These branched-chain  $\alpha$ -keto acid dehydrogenase complexes are similar to pyruvate and  $\alpha$ -ketoglutarate dehydrogenase complexes. Any microorganism that possesses brFAs and/or grows on branched-chain amino acids can be used as a source to isolate bkd genes for expression in host cells, for example, *E. coli*. Furthermore, *E. coli* has the E3 component as part of its pyruvate dehydrogenase complex (lpd, EC 1.8.1.4, GenBank accession NP\_414658). Thus, it may be sufficient to express only the E1  $\alpha/\beta$  and E2 bkd genes. Table 2 lists non-limiting examples of bkd genes from several microorganisms that can be recombinantly introduced and expressed in a host cell to provide branched-chain acyl-CoA precursors.

TABLE 2

Bkd genes from selected microorganisms		
Organism	Gene	GenBank Accession #
<i>Streptomyces coelicolor</i>	bkdA1 (E1 $\alpha$ )	NP_628006
	bkdB1 (E1 $\beta$ )	NP_628005
	bkdc1 (E2)	NP_638004
	bkdA2 (E1 $\alpha$ )	NP_733618
<i>Streptomyces avermitilis</i>	bkdB2 (E1 $\beta$ )	NP_628019
	bkdC2 (E2)	NP_628018
	bkdA (E1a)	BAC72074
	bkdB (E1b)	BAC72075
<i>Streptomyces avermitilis</i>	bkdC (E2)	BAC72076
	bkdF (E1 $\alpha$ )	BAC72088
	bkdG (E1 $\beta$ )	BAC72089
	bkdH (E2)	BAC72090
<i>Bacillus subtilis</i>	bkdAA (E1 $\alpha$ )	NP_390288
	bkdAB (E1 $\beta$ )	NP_390289
	bkdE (E2)	NP_390288
<i>Pseudomonas putida</i>	bkdA1 (E1 $\alpha$ )	AAA65614
	bkdA2 (E1 $\beta$ )	AAA65615
	bkdC (E2)	AAA65617

In another example, isobutyryl-CoA can be made in a host cell, for example in *E. coli*, through the coexpression of a crotonyl-CoA reductase (Ccr, EC 1.6.5.5, 1.1.1.1) and isobutyryl-CoA mutase (large subunit IcmA, EC 5.4.99.2; small subunit IcmB, EC 5.4.99.2) (Han and Reynolds, *J. Bacteriol.* 179:5157, 1997). Crotonyl-CoA is an intermediate in fatty acid biosynthesis in *E. coli* and other microorganisms. Non-limiting examples of ccr and icm genes from selected microorganisms are listed in Table 3.

TABLE 3

Ccr and icm genes from selected microorganisms		
Organism	Gene	GenBank Accession #
<i>Streptomyces coelicolor</i>	ccr	NP_630556
	icmA	NP_629554
	icmB	NP_630904
	crr	AAD53915
<i>Streptomyces cinnamonensis</i>	icmA	AAC08713
	icmB	AJ246005

In addition to expression of the bkd genes, the initiation of brFA biosynthesis utilizes  $\beta$ -ketoacyl-acyl-carrier-protein synthase III (FabH, EC 2.3.1.41) with specificity for branched chain acyl-CoAs (Li et al., *J. Bacteriol.* 187:3795-3799, 2005). Non-limiting examples of such FabH enzymes are listed in Table 4. fabH genes that are involved in fatty acid biosynthesis of any brFA-containing microorganism can be expressed in a host cell. The Bkd and FabH enzymes from host cells that do not naturally make brFA may not support brFA production. Therefore, bkd and fabH can be expressed recombinantly. Vectors containing the bkd and fabH genes can be inserted into such a host cell. Similarly, the endogenous level of Bkd and FabH production may not be sufficient to produce brFA. In this case, they can be overexpressed. Additionally, other components of the fatty acid biosynthesis pathway can be expressed or overexpressed, such as acyl carrier proteins (ACPs) and  $\beta$ -ketoacyl-acyl-carrier-protein synthase II (fabF, EC 2.3.1.41) (non-limiting examples of candidates are listed in Table 4). In addition to expressing these genes, some genes in the endogenous fatty acid biosynthesis pathway can be attenuated in the host cell (e.g., the *E. coli* genes fabH (GenBank accession #NP\_415609) and/or fabF (GenBank accession #NP\_415613)).

TABLE 4

FabH, ACP and fabF genes from selected microorganisms with brFAs		
Organism	Gene	GenBank Accession #
<i>Streptomyces coelicolor</i>	fabH1	NP_626634
	acp	NP_626635
	fabF	NP_626636
	fabH3	NP_823466
<i>Streptomyces avermitilis</i>	fabC3 (acp)	NP_823467
	fabF	NP_823468
	fabH_A	NP_389015
	fabH_B	NP_388898
<i>Bacillus subtilis</i>	acp	NP_389474
	fabF	NP_389016
	fabH	ZP_01643059
	fabF	ZP_01643063
<i>Stenotrophomonas maltophilia</i>	SmalDRAFT_0818 (fabH)	ZP_01643064
	SmalDRAFT_0821 (acp)	ZP_01643064
	SmalDRAFT_0822 (fabF)	ZP_01643064
	Legionella pneumophila	
<i>Legionella pneumophila</i>	fabH	YP_123672
	acp	YP_123675
	fabF	YP_123676

## Formation of Cyclic Fatty Alcohols

Cyclic fatty alcohols can be produced from cyclic fatty aldehydes using cyclic fatty acid derivatives as substrates for a fatty aldehyde biosynthetic polypeptide described herein. To produce cyclic fatty acid derivative substrates, genes that provide cyclic precursors (e.g., the ans, chc, and plm gene families) can be introduced into the host cell and expressed to allow initiation of fatty acid biosynthesis from cyclic precursors. For example, to convert a host cell, such as *E. coli*, into one capable of synthesizing  $\omega$ -cyclic fatty acids (cyFA), a gene that provides the cyclic precursor cyclohexylcarbonyl-CoA (CHC-CoA) (Cropp et al., *Nature Biotech.* 18:980-983, 2000) can be introduced and expressed in the host cell. Non-limiting examples of genes that provide CHC-CoA in *E. coli* include: ansJ, ansK, ansL, chcA, and ansM from the ansatrienin gene cluster of *Streptomyces collinus* (Chen et al., *Eur. J. Biochem.* 261: 98-107, 1999) or plmJ, plmK, plmL, chcA, and plmM from the phoslactomycin B gene cluster of *Streptomyces* sp. HK803 (Palaniappan et al., *J. Biol. Chem.* 278: 35552-35557, 2003) together with the chcB gene (Patton et al., *Biochem.* 39:7595-7604, 2000) from *S. collinus*, *S. avermitilis*, or *S. coelicolor* (see Table 5). The genes listed in Table 4 can then be expressed to allow initiation and elongation of  $\omega$ -cyclic fatty acids. Alternatively, the homologous genes can be isolated from microorganisms that make cyFAs and expressed in a host cell (e.g., *E. coli*).

TABLE 5

Genes for the synthesis of CHC-CoA		
Organism	Gene	GenBank Accession #
<i>Streptomyces collinus</i>	ansJK	U72144*
	ansL	
	chcA	
	ansM	
<i>Streptomyces</i> sp. HK803	cheB	AF268489
	plmJK	AAQ84158
	plmL	AAQ84159
	chcA	AAQ84160
<i>Streptomyces coelicolor</i>	plmM	AAQ84161
	chcB/caiD	NP_629292
<i>Streptomyces avermitilis</i>	chcB/caiD	NP_629292

\*Only chcA is annotated in GenBank entry U72144, ansJKL are according to Chen et al. (*Eur. J. Biochem.* 261: 98-107, 1999).

The genes listed in Table 4 (fabH, acp, and fabF) allow initiation and elongation of  $\omega$ -cyclic fatty acids because they have broad substrate specificity. If the coexpression of any of these genes with the genes listed in Table 5 does not yield cyFA, then fabH, acp, and/or fabF homologs from microorganisms that make cyFAs (e.g., those listed in Table 6) can be isolated (e.g., by using degenerate PCR primers or heterologous DNA sequence probes) and coexpressed.

TABLE 6

Non-limiting examples of microorganisms that contain $\omega$ -cyclic fatty acids	
Organism	Reference
<i>Curtobacterium pusillum</i>	ATCC19096
<i>Alicyclobacillus acidoterrestris</i>	ATCC49025
<i>Alicyclobacillus acidocaldarius</i>	ATCC27009
<i>Alicyclobacillus cycloheptanicus</i> *	Moore, <i>J. Org. Chem.</i> 62: pp. 2173, 1997.

\*Uses cycloheptylcarbonyl-CoA and not cyclohexylcarbonyl-CoA as precursor for cyFA biosynthesis.

## Fatty Alcohol Saturation Levels

The degree of saturation in fatty acids (which can then be converted into fatty aldehydes and then fatty alcohols as

described herein) can be controlled by regulating the degree of saturation of fatty acid intermediates. For example, the sfa, gns, and fab families of genes can be expressed, overexpressed, or expressed at reduced levels, to control the saturation of fatty acids. FIG. 9 lists non-limiting examples of genes in these gene families that may be used in the methods and host cells described herein.

For example, host cells can be engineered to produce unsaturated fatty acids by engineering the production host to 10 overexpress fabB or by growing the production host at low temperatures (e.g., less than 37° C.). FabB has preference to cis- $\delta$ 3-decenoyl-ACP and results in unsaturated fatty acid production in *E. coli*. Overexpression of fabB results in the 15 production of a significant percentage of unsaturated fatty acids (de Mendoza et al., *J. Biol. Chem.* 258:2098-2101, 1983). The gene fabB may be inserted into and expressed in host cells not naturally having the gene. These unsaturated fatty acids can then be used as intermediates in host cells that are engineered to produce fatty acid derivatives, such as fatty 20 aldehydes.

In other instances, a repressor of fatty acid biosynthesis, for 25 example, fabR (GenBank accession NP\_418398), can be deleted, which will also result in increased unsaturated fatty acid production in *E. coli* (Zhang et al., *J. Biol. Chem.* 277: 15558, 2002). Similar deletions may be made in other host 30 cells. A further increase in unsaturated fatty acids may be achieved, for example, by overexpressing fabM (trans-2, cis-3-decenoyl-ACP isomerase, GenBank accession DAA05501) and controlled expression of fabK (trans-2-enoyl-ACP reductase II, GenBank accession NP\_357969) from *Streptococcus pneumoniae* (Marrakchi et al., *J. Biol. Chem.* 277: 44809, 2002), while deleting *E. coli* fabI (trans-2-enoyl-ACP reductase, GenBank accession NP\_415804). In some examples, the endogenous fabF gene can be attenuated, thus increasing the percentage of palmitoleate (C16:1) 35 produced.

In yet other examples, host cells can be engineered to 40 produce saturated fatty acids by reducing the expression of an sfa, gns, and/or fab gene.

In some instances, a host cell can be engineered to express 45 an attenuated level of a dehydratase/isomerase and/or a ketoacyl-ACP synthase. For example, a host cell can be engineered to express a decreased level of fabA and/or fabB. In some instances, the host cell can be grown in the presence of unsaturated fatty acids. In other instances, the host cell can be further 50 engineered to express or overexpress a gene encoding a desaturase enzyme. One nonlimiting example of a desaturase is *B. subtilis* DesA (AF037430). Other genes encoding desaturase enzymes are known in the art and can be used in the host cells and methods described herein, such as desaturases that use acyl-ACP, such as hexadecanoyl-ACP or octadecanoyl-ACP. The saturated fatty acids can be used to produce fatty acid derivatives, such as fatty aldehydes, and subsequently saturated fatty alcohols, as described herein.

## 55 Production of Fatty Alcohols

A fatty aldehyde described herein can be converted into a fatty alcohol by an alcohol dehydrogenase. In some examples, a gene encoding a fatty aldehyde biosynthetic polypeptide described herein can be expressed in a host cell 60 that expresses an endogenous alcohol dehydrogenase capable of converting a fatty aldehyde produced by the fatty aldehyde biosynthetic polypeptide into a corresponding fatty alcohol. In other instances, a gene encoding a fatty alcohol biosynthetic polypeptide described herein, such as an amino acid sequence listed in FIG. 10 or a variant thereof, can be 65 expressed in a host cell. Exemplary fatty alcohol biosynthetic genes include, but are not limited to, AlrA of *Acenitobacter*

sp. M-1 or AlrA homologs; and endogenous *E. coli* alcohol dehydrogenases such as DkgA (NP\_417485), DkgB (NP\_414743), YjgB, (AAC77226), YdjL (AAC74846), YdjJ (NP\_416288), AdhP (NP\_415995), YhdH (NP\_417719), YahK (NP\_414859), YphC (AAC75598), and YqhD (Q46856). In other instances, a gene encoding a fatty alcohol biosynthetic polypeptide can be co-expressed in a host cell with a gene encoding a fatty aldehyde biosynthetic polypeptide described herein.

#### Genetic Engineering of Host Cells to Produce Fatty Alcohols

Various host cells can be used to produce fatty alcohols, as described herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a gene encoding a polypeptide described herein (e.g., a fatty aldehyde biosynthetic polypeptide and/or a fatty alcohol biosynthetic polypeptide) can be expressed in bacterial cells (such as *E. coli*), insect cells, yeast, or mammalian cells (such as Chinese hamster ovary cells (CHO) cells, COS cells, VERO cells, BHK cells, HeLa cells, Cv1 cells, MDCK cells, 293 cells, 3T3 cells, or PC12 cells). Other exemplary host cells include cells from the members of the genus *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*. Yet other exemplary host cells can be a *Bacillus lenthus* cell, a *Bacillus brevis* cell, a *Bacillus stearothermophilus* cell, a *Bacillus licheniformis* cell, a *Bacillus alkalophilus* cell, a *Bacillus coagulans* cell, a *Bacillus circulans* cell, a *Bacillus pumilis* cell, a *Bacillus thuringiensis* cell, a *Bacillus clausii* cell, a *Bacillus megaterium* cell, a *Bacillus subtilis* cell, a *Bacillus amyloliquefaciens* cell, a *Trichoderma koningii* cell, a *Trichoderma viride* cell, a *Trichoderma reesei* cell, a *Trichoderma longibrachiatum* cell, an *Aspergillus awamori* cell, an *Aspergillus fumigatus* cell, an *Aspergillus foetidus* cell, an *Aspergillus nidulans* cell, an *Aspergillus niger* cell, an *Aspergillus oryzae* cell, a *Humicola insolens* cell, a *Humicola lanuginose* cell, a *Rhizomucor miehei* cell, a *Mucor michei* cell, a *Streptomyces lividans* cell, a *Streptomyces murinus* cell, or an *Actinomycetes* cell. Other host cells are cyanobacterial host cells.

In a preferred embodiment, the host cell is an *E. coli* cell, a *Saccharomyces cerevisiae* cell, or a *Bacillus subtilis* cell. In a more preferred embodiment, the host cell is from *E. coli* strains B, C, K, or W. Other suitable host cells are known to those skilled in the art.

Additional host cells that can be used in the methods described herein are described in WO2009/111513 and WO2009/111672.

Various methods well known in the art can be used to genetically engineer host cells to produce fatty alcohols. The methods can include the use of vectors, preferably expression vectors, containing a nucleic acid encoding a fatty aldehyde biosynthetic polypeptide and/or a fatty alcohol biosynthetic polypeptide described herein, polypeptide variant, or a fragment thereof. Those skilled in the art will appreciate a variety of viral vectors (for example, retroviral vectors, lentiviral vectors, adenoviral vectors, and adeno-associated viral vectors) and non-viral vectors can be used in the methods described herein.

The recombinant expression vectors described herein include a nucleic acid described herein in a form suitable for expression of the nucleic acid in a host cell. The recombinant expression vectors can include one or more control sequences, selected on the basis of the host cell to be used for expression. The control sequence is operably linked to the

nucleic acid sequence to be expressed. Such control sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Control sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cells and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors described herein can be introduced into host cells to produce polypeptides, including fusion polypeptides, encoded by the nucleic acids as described herein.

Recombinant expression vectors can be designed for expression of a gene encoding a fatty aldehyde biosynthetic polypeptide (or variant) and/or a gene encoding a fatty alcohol biosynthetic polypeptide in prokaryotic or eukaryotic cells (e.g., bacterial cells, such as *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells). Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990).

Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example, by using T7 promoter regulatory sequences and T7 polymerase.

Expression of genes encoding polypeptides in prokaryotes, for example, *E. coli*, is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: (1) to increase expression of the recombinant polypeptide; (2) to increase the solubility of the recombinant polypeptide; and (3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide. This enables separation of the recombinant polypeptide from the fusion moiety after purification of the fusion polypeptide. Examples of such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin, and enterokinase. Exemplary fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith et al., *Gene* (1988) 67:31-40), pMAL (New England Biolabs, Beverly, Mass.), and pRITS (Pharmacia, Piscataway, N.J.), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant polypeptide.

Examples of inducible, non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* (1988) 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant polypeptide expression is to express the polypeptide in a host cell with an impaired capacity to proteolytically cleave the recombinant

polypeptide (see Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Another strategy is to alter the nucleic acid sequence to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the host cell (Wada et al., *Nucleic Acids Res.* (1992) 20:2111-2118). Such alteration of nucleic acid sequences can be carried out by standard DNA synthesis techniques.

In another embodiment, the host cell is a yeast cell. In this embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYEpSec1 (Baldari et al., *EMBO J.* (1987) 6:229-234), pMFa (Kurjan et al., *Cell* (1982) 30:933-943), pJRY88 (Schultz et al., *Gene* (1987) 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, a polypeptide described herein can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf9 cells) include, for example, the pAc series (Smith et al., *Mol. Cell. Biol.* (1983) 3:2156-2165) and the pVL series (Lucklow et al., *Virology* (1989) 170:31-39).

In yet another embodiment, the nucleic acids described herein can be expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, *Nature* (1987) 329:840) and pMT2PC (Kaufman et al., *EMBO J.* (1987) 6:187-195). When used in mammalian cells, the expression vector's control functions can be provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus, and Simian Virus 40. Other suitable expression systems for both prokaryotic and eukaryotic cells are described in chapters 16 and 17 of Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Vectors can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in, for example, Sambrook et al. (*supra*).

For stable transformation of bacterial cells, it is known that, depending upon the expression vector and transformation technique used, only a small fraction of cells will take-up and replicate the expression vector. In order to identify and select these transformants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Selectable markers include those that confer resistance to drugs, such as ampicillin, kanamycin, chloramphenicol, or tetracycline. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection

technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin, and methotrexate. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

#### 15 Transport Proteins

Transport proteins can export polypeptides and organic compounds (e.g., fatty alcohols) out of a host cell. Many transport and efflux proteins serve to excrete a wide variety of compounds and can be naturally modified to be selective for particular types of hydrocarbons.

Non-limiting examples of suitable transport proteins are ATP-Binding Cassette (ABC) transport proteins, efflux proteins, and fatty acid transporter proteins (FATP). Additional non-limiting examples of suitable transport proteins include the ABC transport proteins from organisms such as *Cae-norhabditis elegans*, *Arabidopsis thaliana*, *Alkaligenes eutrophus*, and *Rhodococcus erythropolis*. Exemplary ABC transport proteins that can be used are listed in FIG. 9 (e.g., CER5, AtMRP5, AmiS2, and AtPGP1). Host cells can also be chosen for their endogenous ability to secrete organic compounds. The efficiency of organic compound production and secretion into the host cell environment (e.g., culture medium, fermentation broth) can be expressed as a ratio of intracellular product to extracellular product. In some examples, the ratio can be about 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4, or 1:5.

#### 16 Fermentation

The production and isolation of fatty alcohols can be enhanced by employing beneficial fermentation techniques.

One method for maximizing production while reducing costs is increasing the percentage of the carbon source that is converted to hydrocarbon products.

During normal cellular lifecycles, carbon is used in cellular functions, such as producing lipids, saccharides, proteins, organic acids, and nucleic acids. Reducing the amount of carbon necessary for growth-related activities can increase the efficiency of carbon source conversion to product. This can be achieved by, for example, first growing host cells to a desired density (for example, a density achieved at the peak of the log phase of growth). At such a point, replication checkpoint genes can be harnessed to stop the growth of cells. Specifically, quorum sensing mechanisms (reviewed in Camilli et al., *Science* 311:1113, 2006; Venturi *FEMS Microbiol. Rev.* 30:274-291, 2006; and Reading et al., *FEMS Microbiol. Lett.* 254:1-11, 2006) can be used to activate checkpoint genes, such as p53, p21, or other checkpoint genes.

Genes that can be activated to stop cell replication and growth in *E. coli* include umuDC genes. The overexpression of umuDC genes stops the progression from stationary phase to exponential growth (Murli et al., *J. of Bact.* 182:1127, 2000). UmuC is a DNA polymerase that can carry out trans-lesion synthesis over non-coding lesions—the mechanistic basis of most UV and chemical mutagenesis. The umuDC gene products are involved in the process of translesion synthesis and also serve as a DNA sequence damage checkpoint. The umuDC gene products include UmuC, UmuD, umuD', UmuD'2C, UmuD'2, and UmuD2. Simultaneously, product-

53

producing genes can be activated, thus minimizing the need for replication and maintenance pathways to be used while a fatty aldehyde is being made. Host cells can also be engineered to express *umuC* and *umuD* from *E. coli* in pBAD24 under the *prpBCDE* promoter system through de novo synthesis of this gene with the appropriate end-product production genes.

The percentage of input carbons converted to fatty alcohols can be a cost driver. The more efficient the process is (i.e., the higher the percentage of input carbons converted to fatty alcohols), the less expensive the process will be. For oxygen-containing carbon sources (e.g., glucose and other carbohydrate based sources), the oxygen must be released in the form of carbon dioxide. For every 2 oxygen atoms released, a carbon atom is also released leading to a maximal theoretical metabolic efficiency of approximately 34% (w/w) (for fatty acid derived products). This figure, however, changes for other organic compounds and carbon sources. Typical efficiencies in the literature are approximately less than 5%. Host cells engineered to produce fatty alcohols can have greater than about 1, 3, 5, 10, 15, 20, 25, and 30% efficiency. In one example, host cells can exhibit an efficiency of about 10% to about 25%. In other examples, such host cells can exhibit an efficiency of about 25% to about 30%. In other examples, host cells can exhibit greater than 30% efficiency.

The host cell can be additionally engineered to express recombinant cellulosomes, such as those described in PCT application number PCT/US2007/003736. These cellulosomes can allow the host cell to use cellulosic material as a carbon source. For example, the host cell can be additionally engineered to express invertases (EC 3.2.1.26) so that sucrose can be used as a carbon source. Similarly, the host cell can be engineered using the teachings described in U.S. Pat. Nos. 5,000,000; 5,028,539; 5,424,202; 5,482,846; and 5,602,030; so that the host cell can assimilate carbon efficiently and use cellulosic materials as carbon sources.

In one example, the fermentation chamber can enclose a fermentation that is undergoing a continuous reduction. In this instance, a stable reductive environment can be created. The electron balance can be maintained by the release of carbon dioxide (in gaseous form). Efforts to augment the NAD/H and NADP/H balance can also facilitate in stabilizing the electron balance. The availability of intracellular NADPH can also be enhanced by engineering the host cell to express an NADH:NADPH transhydrogenase. The expression of one or more NADH:NADPH transhydrogenases converts the NADH produced in glycolysis to NADPH, which can enhance the production of fatty alcohols.

For small scale production, the engineered host cells can be grown in batches of, for example, about 100 mL, 500 mL, 1 L, 2 L, 5 L, or 10 L; fermented; and induced to express desired fatty aldehyde biosynthetic genes and/or an alcohol dehydrogenase genes based on the specific genes encoded in the appropriate plasmids. For large scale production, the engineered host cells can be grown in batches of about 10 L, 100 L, 1000 L, 10,000 L, 100,000 L, 1,000,000 L or larger; fermented; and induced to express desired fatty aldehyde biosynthetic genes and/or alcohol dehydrogenase genes based on the specific genes encoded in the appropriate plasmids or incorporated into the host cell's genome.

For example, a suitable production host, such as *E. coli* cells, harboring plasmids containing the desired genes or having the genes integrated in its chromosome can be incubated in a suitable reactor, for example a 1 L reactor, for 20 hours at 37° C. in M9 medium supplemented with 2% glucose, carbenicillin, and chloramphenicol. When the OD<sub>600</sub> of the culture reaches 0.9, the production host can be induced

54

with IPTG alcohol. After incubation, the spent media can be extracted and the organic phase can be examined for the presence of fatty alcohols using GC-MS.

In some instances, after the first hour of induction, aliquots of no more than about 10% of the total cell volume can be removed each hour and allowed to sit without agitation to allow the fatty alcohols to rise to the surface and undergo a spontaneous phase separation or precipitation. The fatty alcohol component can then be collected, and the aqueous phase returned to the reaction chamber. The reaction chamber can be operated continuously. When the OD<sub>600</sub> drops below 0.6, the cells can be replaced with a new batch grown from a seed culture.

#### Producing Fatty Alcohols Using Cell-Free Methods

In some methods described herein, a fatty alcohol can be produced using a purified polypeptide (e.g., a fatty alcohol biosynthetic polypeptide) described herein and a substrate (e.g., fatty aldehyde), produced, for example, by a method described herein. For example, a host cell can be engineered to express a fatty alcohol biosynthetic polypeptide or variant as described herein. The host cell can be cultured under conditions suitable to allow expression of the polypeptide. Cell free extracts can then be generated using known methods. For example, the host cells can be lysed using detergents or by sonication. The expressed polypeptides can be purified using known methods. After obtaining the cell free extracts, substrates described herein can be added to the cell free extracts and maintained under conditions to allow conversion of the substrates (e.g., fatty aldehydes) to fatty alcohols. The fatty alcohols can then be separated and purified using known techniques.

In some instances, a fatty aldehyde described herein can be converted into a fatty alcohol by contacting the fatty aldehyde with a fatty alcohol biosynthetic polypeptide listed in FIG. 10, or a variant thereof.

#### Post-Production Processing

The fatty alcohols produced during fermentation can be separated from the fermentation media. Any known technique for separating fatty alcohols from aqueous media can be used. One exemplary separation process is a two phase (bi-phasic) separation process. This process involves fermenting the genetically engineered host cells under conditions sufficient to produce a fatty alcohols, allowing the fatty alcohol to collect in an organic phase, and separating the organic phase from the aqueous fermentation broth. This method can be practiced in both a batch and continuous fermentation processes.

Bi-phasic separation uses the relative immiscibility of fatty alcohols to facilitate separation. Immiscible refers to the relative inability of a compound to dissolve in water and is defined by the compound's partition coefficient. One of ordinary skill in the art will appreciate that by choosing a fermentation broth and organic phase, such that the fatty alcohol being produced has a high log P value, the fatty alcohol can separate into the organic phase, even at very low concentrations, in the fermentation vessel.

The fatty alcohols produced by the methods described herein can be relatively immiscible in the fermentation broth, as well as in the cytoplasm. Therefore, the fatty alcohol can collect in an organic phase either intracellularly or extracellularly. The collection of the products in the organic phase can lessen the impact of the fatty alcohol on cellular function and can allow the host cell to produce more product.

The methods described herein can result in the production of homogeneous compounds wherein at least about 60%, 70%, 80%, 90%, or 95% of the fatty alcohols produced will have carbon chain lengths that vary by less than about 6

carbons, less than about 4 carbons, or less than about 2 carbons. These compounds can also be produced with a relatively uniform degree of saturation. These compounds can be used directly as fuels, fuel additives, starting materials for production of other chemical compounds (e.g., polymers, surfactants, plastics, textiles, solvents, adhesives, etc.), or personal care additives. These compounds can also be used as feedstock for subsequent reactions, for example, hydrogenation, catalytic cracking (e.g., via hydrogenation, pyrolysis, or both), to make other products.

In some embodiments, the fatty alcohols produced using methods described herein can contain between about 50% and about 90% carbon; or between about 5% and about 25% hydrogen. In other embodiments, the fatty alcohols produced using methods described herein can contain between about 65% and about 85% carbon; or between about 10% and about 15% hydrogen.

#### Surfactant and Detergent Compositions and Bioproducts

The fatty alcohols described herein can be used as or converted into a surfactant or detergent composition. One of ordinary skill in the art will appreciate that, depending upon the intended purpose of the surfactant or detergent, different fatty alcohols can be produced and used. For example, when the fatty alcohols described herein are used as a feedstock for surfactant or detergent production, one of ordinary skill in the art will appreciate that the characteristics of the fatty alcohol feedstock will affect the characteristics of the surfactant or detergent produced. Hence, the characteristics of the surfactant or detergent product can be selected for by producing particular fatty alcohols for use as a feedstock.

Bioproducts (e.g., fatty alcohols) comprising biologically produced organic compounds, particularly fatty alcohols biologically produced using the fatty acid biosynthetic pathway, have not been produced from renewable sources and, as such, are new compositions of matter. These new bioproducts can be distinguished from organic compounds derived from petrochemical carbon on the basis of dual carbon-isotopic fingerprinting or  $^{14}\text{C}$  dating. Additionally, the specific source of biosourced carbon (e.g., glucose vs. glycerol) can be determined by dual carbon-isotopic fingerprinting (see, e.g., U.S. Pat. No. 7,169,588, which is herein incorporated by reference).

The ability to distinguish bioproducts from petroleum based organic compounds is beneficial in tracking these materials in commerce. For example, organic compounds or chemicals comprising both biologically based and petroleum based carbon isotope profiles may be distinguished from organic compounds and chemicals made only of petroleum based materials. Hence, the instant materials may be followed in commerce on the basis of their unique carbon isotope profile.

Bioproducts can be distinguished from petroleum based organic compounds by comparing the stable carbon isotope ratio ( $^{13}\text{C}/^{12}\text{C}$ ) in each fuel. The  $^{13}\text{C}/^{12}\text{C}$  ratio in a given bioproduct is a consequence of the  $^{13}\text{C}/^{12}\text{C}$  ratio in atmospheric carbon dioxide at the time the carbon dioxide is fixed. It also reflects the precise metabolic pathway. Regional variations also occur. Petroleum,  $\text{C}_3$  plants (the broadleaf),  $\text{C}_4$  plants (the grasses), and marine carbonates all show significant differences in  $^{13}\text{C}/^{12}\text{C}$  and the corresponding  $\delta^{13}\text{C}$  values. Furthermore, lipid matter of  $\text{C}_3$  and  $\text{C}_4$  plants analyze differently than materials derived from the carbohydrate components of the same plants as a consequence of the metabolic pathway.

Within the precision of measurement,  $^{13}\text{C}$  shows large variations due to isotopic fractionation effects, the most significant of which for bioproducts is the photosynthetic

mechanism. The major cause of differences in the carbon isotope ratio in plants is closely associated with differences in the pathway of photosynthetic carbon metabolism in the plants, particularly the reaction occurring during the primary carboxylation (i.e., the initial fixation of atmospheric  $\text{CO}_2$ ). Two large classes of vegetation are those that incorporate the " $\text{C}_3$ " (or Calvin-Benson) photosynthetic cycle and those that incorporate the " $\text{C}_4$ " (or Hatch-Slack) photosynthetic cycle.

In  $\text{C}_3$  plants, the primary  $\text{CO}_2$  fixation or carboxylation reaction involves the enzyme ribulose-1,5-diphosphate carboxylase, and the first stable product is a 3-carbon compound.  $\text{C}_3$  plants, such as hardwoods and conifers, are dominant in the temperate climate zones.

In  $\text{C}_4$  plants, an additional carboxylation reaction involving another enzyme, phosphoenol-pyruvate carboxylase, is the primary carboxylation reaction. The first stable carbon compound is a 4-carbon acid that is subsequently decarboxylated. The  $\text{CO}_2$  thus released is refixed by the  $\text{C}_3$  cycle. Examples of  $\text{C}_4$  plants are tropical grasses, corn, and sugar cane.

Both  $\text{C}_4$  and  $\text{C}_3$  plants exhibit a range of  $^{13}\text{C}/^{12}\text{C}$  isotopic ratios, but typical values are about -7 to about -13 per mil for  $\text{C}_4$  plants and about -19 to about -27 per mil for  $\text{C}_3$  plants (see, e.g., Stuiver et al., *Radiocarbon* 19:355, 1977). Coal and petroleum fall generally in this latter range. The  $^{13}\text{C}$  measurement scale was originally defined by a zero set by Pee Dee Belemnite (PDB) limestone, where values are given in parts per thousand deviations from this material. The " $\delta^{13}\text{C}$ " values are expressed in parts per thousand (per mil), abbreviated, ‰, and are calculated as follows:

$$\delta^{13}\text{C}(\text{‰}) = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} - (^{13}\text{C}/^{12}\text{C})_{\text{standard}}] / [(^{13}\text{C}/^{12}\text{C})_{\text{standard}} \times 1000]$$

Since the PDB reference material (RM) has been exhausted, a series of alternative RMs have been developed in cooperation with the IAEA, USGS, NIST, and other selected international isotope laboratories. Notations for the per mil deviations from PDB is  $\delta^{13}\text{C}$ . Measurements are made on  $\text{CO}_2$  by high precision stable ratio mass spectrometry (IRMS) on molecular ions of masses 44, 45, and 46.

The compositions described herein include bioproducts produced by any of the methods described herein. Specifically, the bioproduct can have a  $\delta^{13}\text{C}$  of about -28 or greater, about -27 or greater, -20 or greater, -18 or greater, -15 or greater, -13 or greater, -10 or greater, or -8 or greater. For example, the bioproduct can have a  $\delta^{13}\text{C}$  of about -30 to about -15, about -27 to about -19, about -25 to about -21, about -15 to about -5, about -13 to about -7, or about -13 to about -10. In other instances, the bioproduct can have a  $\delta^{13}\text{C}$  of about -10, -11, -12, or -12.3.

Bioproducts can also be distinguished from petroleum based organic compounds by comparing the amount of  $^{14}\text{C}$  in each compound. Because  $^{14}\text{C}$  has a nuclear half life of 5730 years, petroleum based fuels containing "older" carbon can be distinguished from bioproducts which contain "newer" carbon (see, e.g., Currie, "Source Apportionment of Atmospheric Particles", *Characterization of Environmental Particles*, J. Buffel and H. P. van Leeuwen, Eds., 1 of Vol. I of the IUPAC Environmental Analytical Chemistry Series (Lewis Publishers, Inc) (1992) 3-74).

The basic assumption in radiocarbon dating is that the constancy of  $^{14}\text{C}$  concentration in the atmosphere leads to the constancy of  $^{14}\text{C}$  in living organisms. However, because of atmospheric nuclear testing since 1950 and the burning of fossil fuel since 1850,  $^{14}\text{C}$  has acquired a second, geochemical time characteristic. Its concentration in atmospheric  $\text{CO}_2$ , and hence in the living biosphere, approximately doubled at the peak of nuclear testing, in the mid-1960s. It has since been

57

gradually returning to the steady-state cosmogenic (atmospheric) baseline isotope rate ( $^{14}\text{C}/^{12}\text{C}$ ) of about  $1.2 \times 10^{-12}$ , with an approximate relaxation “half-life” of 7-10 years. (This latter half-life must not be taken literally; rather, one must use the detailed atmospheric nuclear input/decay function to trace the variation of atmospheric and biospheric  $^{14}\text{C}$  since the onset of the nuclear age.)

It is this latter biospheric  $^{14}\text{C}$  time characteristic that holds out the promise of annual dating of recent biospheric carbon.  $^{14}\text{C}$  can be measured by accelerator mass spectrometry (AMS), with results given in units of “fraction of modern carbon” ( $f_M$ ).  $f_M$  is defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C. As used herein, “fraction of modern carbon” or “ $f_M$ ” has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOxI and HOxII, respectively. The fundamental definition relates to 0.95 times the  $^{14}\text{C}/^{12}\text{C}$  isotope ratio HOxI (referenced to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (plant material),  $f_M$  is approximately 1.1.

The compositions described herein include bioproducts that can have an  $f_M$   $^{14}\text{C}$  of at least about 1. For example, the bioproduct can have an  $f_M$   $^{14}\text{C}$  of at least about 1.01, an  $f_M$   $^{14}\text{C}$  of about 1 to about 1.5, an  $f_M$   $^{14}\text{C}$  of about 1.04 to about 1.18, or an  $f_M$   $^{14}\text{C}$  of about 1.111 to about 1.124.

Another measurement of  $^{14}\text{C}$  is known as the percent of modern carbon, pMC. For an archaeologist or geologist using  $^{14}\text{C}$  dates, AD 1950 equals “zero years old”. This also represents 100 pMC. “Bomb carbon” in the atmosphere reached almost twice the normal level in 1963 at the peak of thermonuclear weapons. Its distribution within the atmosphere has been approximated since its appearance, showing values that are greater than 100 pMC for plants and animals living since AD 1950. It has gradually decreased over time with today’s value being near 107.5 pMC. This means that a fresh biomass material, such as corn, would give a  $^{14}\text{C}$  signature near 107.5 pMC. Petroleum based compounds will have a pMC value of zero. Combining fossil carbon with present day carbon will result in a dilution of the present day pMC content. By presuming 107.5 pMC represents the  $^{14}\text{C}$  content of present day biomass materials and 0 pMC represents the  $^{14}\text{C}$  content of petroleum based products, the measured pMC value for that material will reflect the proportions of the two component types. For example, a material derived 100% from present day soybeans would give a radiocarbon signature near 107.5 pMC. If that material was diluted 50% with petroleum based products, it would give a radiocarbon signature of approximately 54 pMC.

A biologically based carbon content is derived by assigning “100%” equal to 107.5 pMC and “0%” equal to 0 pMC. For example, a sample measuring 99 pMC will give an equivalent biologically based carbon content of 93%. This value is referred to as the mean biologically based carbon result and assumes all the components within the analyzed material originated either from present day biological material or petroleum based material.

A bioproduct described herein can have a pMC of at least about 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, or 100. In other instances, a bioproduct described herein can have a pMC of between about 50 and about 100; about 60 and about 100; about 70 and about 100; about 80 and about 100; about 85 and about 100; about 87 and about 98; or about 90 and about 95. In yet other instances, a bioproduct described herein can have a pMC of about 90, 91, 92, 93, 94, or 94.2.

58

Fuel additives are used to enhance the performance of a fuel or engine. For example, fuel additives can be used to alter the freezing/gelling point, cloud point, lubricity, viscosity, oxidative stability, ignition quality, octane level, and/or flash point. In the United States, all fuel additives must be registered with Environmental Protection Agency. The names of fuel additives and the companies that sell the fuel additives are publicly available by contacting the EPA or by viewing the agency’s website. One of ordinary skill in the art will appreciate that the fatty alcohol-based biofuels described herein can be mixed with one or more fuel additives to impart a desired quality.

The fatty alcohol-based surfactants and/or detergents described herein can be mixed with other surfactants and/or detergents well known in the art.

In some examples, the mixture can include at least about 10%, 15%, 20%, 30%, 40%, 50%, or 60% by weight of the fatty alcohol. In other examples, a surfactant or detergent composition can be made that includes at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or 95% of a fatty alcohol that includes a carbon chain that is 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 carbons in length. Such surfactant or detergent compositions can additionally include at least one additive selected from a surfactant; a microemulsion; at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, or 95% of surfactant or detergent from nonmicrobial sources such as plant oils or petroleum.

The invention is further illustrated by the following examples. The examples are provided for illustrative purposes only. They are not to be construed as limiting the scope or content of the invention in any way.

## EXAMPLES

### Example 1

#### Identification of Carboxylic Acid Reductase (CAR) Homologs

The carboxylic acid reductase (CAR) from *Nocardia* sp. strain NRRL 5646 can reduce carboxylic acids (e.g., fatty acids) into their corresponding aldehydes without utilizing separate activating enzymes, such as acyl-CoA synthases (Li et al., *J. Bacteriol.* 179:3482-3487, 1997; He et al., *Appl. Environ. Microbiol.* 70:1874-1881, 2004)).

A BLAST search using the NRRL 5646 CAR amino acid sequence (Genpept accession AAR91681) (SEQ ID NO:16) as the query sequence identified approximately 20 homologous sequences. Three homologs, listed in Table 7, were evaluated for their ability to convert fatty acids into fatty aldehydes *in vivo* when expressed in *E. coli*.

At the nucleotide sequence level, carA (SEQ ID NO:19), carB (SEQ ID NO:21), and fadD9 (SEQ ID NO:17) demonstrated 62.6%, 49.4%, and 60.5% homology, respectively, to the car gene (AY495697) of *Nocardia* sp. NRRL 5646 (SEQ ID NO:15). At the amino acid level, CARA (SEQ ID NO:20), CARB (SEQ ID NO:22), and FadD9 (SEQ ID NO:18) demonstrated 62.4%, 59.1% and 60.7% identity, respectively, to CAR of *Nocardia* sp. NRRL 5646 (SEQ ID NO:16).

TABLE 7

CAR-like Protein and the corresponding coding sequences.			
Genpept accession	Locus_tag	Annotation in GenBank	Gene name
NP_217106	Rv 2590	Probable fatty-acid-CoA ligase (FadD9)	fadD9
ABK75684	MSMEG 2956	NAD dependent epimerase/dehydratase family protein	carA
YP_889972.1	MSMEG 5739	NAD dependent epimerase/dehydratase family protein	carB

## Example 2

## Identification of Alcohol Dehydrogenase Genes

## Reverse Engineering

*E. coli* contains at least one enzyme that catalyzes the reversible oxidoreduction of fatty aldehydes and fatty alcohols (i.e. fatty aldehyde reductase/alcohol dehydrogenase). Reverse engineering was used to identify such fatty aldehyde reductases/fatty alcohol dehydrogenases in *E. coli* MG1655 cells expressing the acyl-ACP reductase YP\_400611 from *Synechococcus elongatus* (Synpcc7942\_1594) (SEQ ID NO:196). Four 3 mL LB cultures were grown overnight at 37° C., and 55 µL of stationary phase cultures were used to inoculate four independent 5.5 mL of LB. Those 5.5 mL cultures were then grown to an OD<sub>600</sub> of 0.8-1.0 and were then used to inoculate a corresponding number of 2 L baffled shakeflasks, each with 500 mL Hu-9 minimal media. 20 hrs after induction the cells were pelleted at 4,000×g for 20 min. The cell pellet was resuspended in 30 mL of 100 mM phosphate buffer at pH 7.2 with 1× Bacterial Protease Arrest (G Biosciences). The cells were lysed in a french press at 15,000 psi with two passes through the instrument. The cell debris was then removed by centrifuging at 10,000×g for 20 mins. The cell lysate was loaded onto two HiTrapQ columns (GE Healthcare) connected in series. The following buffers were used to elute proteins: (A) 50 mM Tris, pH 7.5 and (B) 50 mM Tris, pH 7.5 with 1 M NaCl. A gradient from 0% B to 100% B was run over 5 column volumes at a flow rate of 3 mL/min while 4 mL fractions were collected.

The fractions were assayed for alcohol dehydrogenase activity by taking 190 µL of a protein fraction and adding 5 µL of a 20 mM NADPH (Sigma) solution and 5 µL of a 20 mM dodecanal (Fluka) solution in DMSO. The reactions were incubated at 37° C. for 1 hr. They were then extracted with 100 µL of ethyl acetate and analyzed for dodecanol via GC/MS. Fractions eluting around 350 mM NaCl contained alcohol dehydrogenase activity.

Fractions containing alcohol dehydrogenase activity were pooled and loaded onto a 1 mL ResourceQ column (GE Healthcare). The same conditions used for the HiTrapQ column were used, except 0.5 mL fractions were collected. Protein fractions demonstrating alcohol dehydrogenase activity were then pooled and concentrated using Amicon (Millipore) protein concentrators (10,000 kDa cutoffs) to a volume of 1 mL. The solution was then loaded onto a HiPrep 200 size exclusion column (GE Healthcare). A buffer solution containing 50 mM Tris, pH 7.5, and 150 mM NaCl was run through the column at a rate of 0.3 mL per min. 2 mL fractions were collected. Two protein fractions contained alcohol dehydrogenase activity. These two fractions, plus fractions

before and after these two fractions, were loaded onto a polyacrylamide gel and stained with SimplySafe Commassie stain (Invitrogen).

Comparing the bands in the active and inactive fractions, one protein band appeared in the active fractions that was not seen in the inactive fraction. This protein band was cut from gel and submitted to the Stanford Mass Spectroscopy Facility for LC/MS/MS protein sequencing. One of the proteins identified in this analysis was YahK.

To verify that YahK was indeed an alcohol dehydrogenase, yahK was knocked out in *E. coli* MG1655(DE3, ΔfadD, ΔyjgB) (control strain) (described in Example 4). The yahK knock-out strain MG1655(DE3, ΔfadD, Δyjg, ΔyahK) was constructed using the lambda red system (described in Example 4) with the following primers:

yahK\_F  
(SEQ ID NO: 197)  
20 (CATATCAGGCCTTGCCAAATACACATAGCTAACAGGAGTAAACACAA  
TG)  
and  
yahK\_R  
(SEQ ID NO: 198)  
25 (AATCGCACACTAACAGACTGAAAAAATTAAATAAAACCCCTGTGGTTA  
AC) .

This ΔyahK strain and the control strain, both expressing the acyl-ACP reductase YP\_400611, were cultured under conditions described above. Cell free lysates were made from both strains, and each lysate was assayed for alcohol dehydrogenase activity as discussed above.

The ΔyahK strain did not convert dodecanal to dodecanol, while the wild type strain had this activity. For additional verification, each lysate was run on a HiTrapQ column as described above. The wild type lysate had alcohol dehydrogenase activity in fractions eluting around 350 mM NaCl, while the ΔyahK lysate had no alcohol dehydrogenase activity in this region.

Bioinformatics  
45 It was reasoned that possible alcohol dehydrogenases in *E. coli* were members of four protein families: Zn-dependent alcohol dehydrogenases (Pfam 00107 and 08240), Fe-dependent alcohol dehydrogenases (Pfam 00465), aldo-keto reductases (Pfam 00248) and short-chain dehydrogenases (Pfam 00106) (Pfam=protein family according to “pfam.sanger.ac.uk”). Using the Pfam motifs, all members of these four protein families in *E. coli* were identified (listed in FIG. 10).  
50 From this list, the following 8 candidates were chosen for experimental analysis: yahK, yjgB, adhP, dkgA, dkgB, yhdH, ydjL, and yqhD.

To determine if these genes could reduce fatty aldehydes to fatty alcohols, these 8 genes were cloned into a pET-Duet vector along with *E. coli* ‘tesA. These genes were then transformed into *E. coli* (DE3) MG1655 ΔyjgBΔyahK cells. Next 3 mL overnight starter cultures were grown in LB with carbanecillin (100 mg/L) at 37° C. A control strain lacking a candidate alcohol dehydrogenase was also included in the experiment. 1 mL of each overnight culture was used to inoculate 50 mL of fresh LB with carbanecillin. The cultures were shaken at 37° C. until reaching an OD<sub>600</sub> of 0.8-1. The cultures were then transferred to 18° C., induced with 1 mM IPTG, and shaken overnight.

Cell free lysates were prepared by centrifuging the cultures at 4,000×g for 20 mins. The cultures were then resuspended in 1 mL of Bugbuster (Novagen) and gently shaken at room

**61**

temperature for 5 min. The cell debris was removed by spinning at 15,000×g for 10 min. The resulting lysates were assayed for alcohol dehydrogenase activity by mixing 88 μL of lysate, 2 μL of 40 mM cis-11-hexadecenal in DMSO, and 10 μL of 20 mM NADPH. The samples were incubated at 37° C. for 30 min. and were then extracted with 100 μL of ethyl acetate. The extracts were analyzed using GC/MS.

All proteins showed significantly better conversion of cis-11-hexadecenal to cis-11-hexadecanol as compared with the 'TesA only control (see Table 8). These results were confirmed in assays using dodecanal instead of cis-11-hexadecenal as the substrate (see Table 8).

To investigate how these enzymes contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions, first the yjgB yahK double knock-out strain in MG1655(DE3, ΔfadD) (described above) was tested by transforming it with a plasmid expressing acyl-ACP reductase YP\_400611 and analyzing fatty aldehyde and fatty alcohol titers. The test strain also contained a plasmid expressing a decarbonylase. This double knock-out mutant showed slightly higher fatty aldehyde titers in several experiments (see, e.g., FIG. 11), confirming that these two putative alcohol dehydrogenases contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions (see also Example 4 for similar results from a MG1655(DE3, ΔfadD ΔyjgB) strain). Next, two additional genes, yncB and ydjA, were deleted in the yjgB yahK double mutant. YdjA, which is not a member of the four protein families mentioned above, demonstrated slightly elevated fatty aldehyde levels (see FIG. 11), indicating that it may also contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions.

Additionally, the active fatty alcohol dehydrogenases from Table 8 were also deleted in MG1655 (DE3, ΔfadD, Δyjg, B ΔyahK) and tested as described above. Several of these deletion strains showed slightly elevated fatty aldehyde levels, suggesting that these may also contribute to fatty alcohol dehydrogenase activity in *E. coli* under production conditions (see FIG. 12).

TABLE 8

Overexpression of putative fatty alcohol dehydrogenase genes			
substrate	GC/MS Assay		
	dodecanal	% conversion to corresponding alcohol	NADPH assay initial rate (slope)
Overexpression:			
none	9	12	0.2
YjgB	54	89	24.8
YahK	47	87	28.3
AdhP	52	45	4.1
YdjL	51	23	0.14
YhdH	59	74	13.7
YqhD	55	23	7.3
yafB (dkgB)	52	65	9.4
YqhE (dkgA)	45	50	9.6

**62**

## Example 3

Expression of CAR Homologs and Alcohol Dehydrogenase in *E. coli*

5

## A. CAR Plasmid Construction

Three *E. coli* expression plasmids were constructed to express the genes encoding the CAR homologs listed in Table 7. First, fadD9 was amplified from genomic DNA of *Mycobacterium tuberculosis* H37Rv (obtained from The University of British Columbia, and Vancouver, BC Canada) using the primers fadD9F and FadDR (see Table 9). The PCR product was first cloned into PCR-blunt (Invitrogen) and then released as an NdeI-AvrII fragment. The NdeI-AvrII fragment was then cloned between the NdeI and AvrII sites of pACYCDuet-1 (Novogen) to generate pACYCDuet-1-fadD9.

The carA gene was amplified from the genomic DNA of *Mycobacterium smegmatis* MC2 155 (obtained from the ATCC (ATCC 23037D-5)) using primers CARMCaF and CARMCaR (see Table 9). The carB gene was amplified from the genomic DNA of *Mycobacterium smegmatis* MC2 155 (obtained from the ATCC (ATCC 23037D-5)) using primers CARMCbF and CARMCbR (see Table 9). Each PCR product was first cloned into PCR-blunt and then released as an NdeI-AvrII fragment. Each of the two fragments was then subcloned between the NdeI and AvrII sites of pACYCDuet-1 (Novogen) to generate pACYCDuet-1-carA and pACYCDuet-1-carB.

TABLE 9

## Primers used to amplify genes encoding CAR homologs

fadD9F	cat ATGTCGATCAACGATCAGCGACTGAC (SEQ ID NO: 1)
fadD9R	cctagg TCACAGCAGCCGAGCAGTC (SEQ ID NO: 2)
CARMCaF	cat ATGACGATCGAAACGCG (SEQ ID NO: 3)
CARMCaR	cctagg TTACAGCAATCCGAGCATCT (SEQ ID NO: 4)
CARMCbF	cat ATGACCAGCGATGTTCAC (SEQ ID NO: 5)
CARMCbR	cctagg TCAGATCAGACCGAACTCACG (SEQ ID NO: 6)

## B. Alcohol Dehydrogenase Plasmid Construction

The plasmid pETDuet-1-'tesA-yjgB carries 'tesA and yjgB (a putative alcohol dehydrogenase; GenBank accession number, NP\_418690; GenPept accession number AAC77226) from the *E. coli* K12 strain.

The gene yjgB (GenBank accession number, NP\_418690) was amplified from the genomic DNA of *E. coli* K-12 using the following primers.

55 The yjgB insert was generated by PCR using the following primers:

NcoI YjgB forward:  
(SEQ ID NO: 199)  
aatccTGGCATCGATGATAAAAAGCTATGCCGCAAAG

HindIII YjgB reverse:  
(SEQ ID NO: 200)  
ataaaagctTCAAAATCGGCTTCAACACCGCGG

60 The PCR product was then subcloned into the NcoI and HindIII sites of pETDuet-1-'tesA to generate pETDuet-1-'tesA-yjgB.

The plasmid pETDuet-1-'tesA-alrAadp1 carries 'tesA and alrAadp1 (GenPept accession number CAG70248.1) from *Acinetobacter baylyi* ADP1.

The gene alrAadp1 was amplified from the genomic DNA of *Acinetobacter baylyi* ADP1 by a two-step PCR procedure. The first set of PCR reactions eliminated an internal NcoI site at by 632-636 with the following primer pairs:

ADP1 Alr mut1 reverse:  
 (SEQ ID NO: 201)  
 5'-GACCACTGTATGGCCCCCATAGCTTGGAGCTCATC

ADP1 Alr1 mut1 forward:  
 (SEQ ID NO: 202)  
 5'-GATGAGCTAAAGCTATGGGGCCGATCACGTGGTC

The PCR products were then isolated, purified using the Qiagen gel extraction kit, and used as inputs for a second PCR reaction with the following primers to produce full-length AlrAadp1 with a C→T mutation at position 633:

NcoI ADP1 Alr1 forward:  
 (SEQ ID NO: 203)  
 5'-AATACCATGGCAACAACTAATGTGATTCAATGCTTATGCTGCA

HindIII ADP1 Alr1 reverse:  
 (SEQ ID NO: 204)  
 5'-ATAAAAGCTTTAAAAATCGGCTTAAGTACAATCCGATAAC

The plasmid pETDuet-1-'tesA-alrAadp1 was prepared by inserting the alrAadp1 gene (gene locus-tag="ACIAD3612"), a homolog of *Acinetobacter baylyi* ADP1, into the NcoI and HindIII sites of pETDuet-1-'tesA. B. Evaluation of Fatty Aldehyde and Fatty Alcohol Production

In order to evaluate the affect of carboxylic acid reductases and alcohol dehydrogenases on the production of fatty alcohols, various combinations of the prepared plasmids were transformed in the *E. coli* strain C41 (DE3, ΔfadE) (described in PCT/US08/058,788).

For example, the plasmid pACYCDuet-1-carA, encoding the CAR homolog carA, was co-transformed with pETDuet-1-'tesA-alrAadp1 (see, e.g., FIG. 1).

The plasmid pACYCDuet-1-carB, encoding the CAR homolog carB, was co-transformed with pETDuet-1-'tesA. In addition, pACYCDuet-1-carB was also separately co-transformed with pETDuet-1-'tesA-yjgB and pETDuet-1-'tesA-alrAadp1. As a control, pACYCDuet-1-carB was co-transformed with the empty vector pETDuet-1 (see, e.g., FIG. 1).

The plasmid pACYCDuet-1-fadD9, encoding the CAR homolog fadD9, was co-transformed with pETDuet-1-'tesA. In addition, pACYCDuet-1-fadD9 was also separately co-transformed with pETDuet-1-'tesA-yjgB and pETDuet-1-'tesA-alrAadp1. As a control, pACYCDuet-1-fadD9 was co-transformed with the empty vector pETDuet-1 (see, e.g., FIG. 1).

As an additional control, pETDuet-1-'tesA-yjgB was co-transformed with the empty vector pACYCDuet-1.

The *E. coli* transformants were grown in 3 mL of LB medium supplemented with carbenicillin (100 mg/L) and chloramphenicol (34 mg/L) at 37° C. After overnight growth, 15 μL of culture was transferred into 2 mL of fresh LB medium supplemented with carbenicillin and chloramphenicol. After 3.5 hours of growth, 2 mL of culture were transferred into a 125 mL flask containing 20 mL of M9 medium with 2% glucose and with carbenicillin and chloramphenicol. When the OD<sub>600</sub> of the culture reached 0.9, 1 mM of IPTG

was added to each flask. After 20 hours of growth at 37° C., 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described herein.

The measured retention times were 6.79 minutes for cis-5-dodecen-1-ol, 6.868 minutes for 1-dodecanol, 8.058 minutes for cis-7-tetradecen-1-ol, 8.19 minutes for 1-tetradecanol, 9.208 minutes for cis-9-hexadecen-1-ol, 9.30 minutes for 1-hexadecanol, and 10.209 minutes for cis-11-octadecen-1-ol.

The co-expression of the leaderless tesA and any of the three car genes in *E. coli* resulted in high titers of fatty alcohols and detectable fatty aldehyde production (FIGS. 1, 2, 5).

15 The expression of carA or carB with the leaderless tesA and alrAadp1 resulted in fatty alcohol titers of greater than 700 mg/L and reduced fatty aldehyde production. Likewise, fadD9 co-expressed with the leaderless tesA and alrAadp1 produced over 300 mg/L of fatty alcohol. When expressed without the leaderless tesA, neither carB nor fadD9 produced more than 10 mg/L of fatty alcohols (possibly resulting from the accumulation of free fatty acids in the cell due to endogenous tesA). Taken together, this data indicates that fatty acids are the substrates for these CAR homologs and that 20 overexpression of a thioesterase, such as 'tesA (to release fatty acids from acyl-ACP), achieves significant production of fatty alcohols.

In one fermentation, *E. coli* strain C41 (DE3, AfadE) co-transformed with pACYCDuet-1-carB+pETDuet-1-tesA produced an average of 695 mg/L of fatty alcohols and 120 mg/L of fatty aldehydes. The presence of large amounts of fatty aldehydes is consistent with CAR being an aldehyde-generating, fatty acid reductase (AFAR). This mechanism is different from alcohol-generating fatty acyl-CoA reductases (FAR), represented by JfFAR, and fatty acyl-CoA reductases, represented by Acr1.

The production of fatty alcohols from fatty aldehydes in the *E. coli* strains described above may have been catalyzed by an endogenous alcohol dehydrogenase(s). *E. coli* produces an 30 alcohol dehydrogenase(s) (e.g., yjgB) capable of converting fatty aldehydes of various chain-length into fatty alcohols (Naccarato et al., *Lipids* 9: 419-428 (1974); Reiser et al., *J. Bacteriol.* 179: 2969-2975 (1997); Venkatasubramanian et al., *J. Biol. Chem.* 282:478-485 (2007)).

40 Therefore, alcohols dehydrogenases may also play a role in the fatty alcohol biosynthetic pathway in addition to carboxylic acid reductases. For example, expression of either yjgB or alrAadp1 with carB and the leaderless tesA significantly reduced the accumulation of fatty aldehydes, compared to strains that did not overexpress yjgB or alrAadp1 (FIG. 2).

Following the fermentations where pACYCDuet-1-carB was transformed in *E. coli* strain C41 (DE3, ΔfadE), a white, round, disk-like deposit was observed at the bottom center of the flasks used for fatty alcohol production with recombinant 45 *E. coli* strains. In contrast, no such deposits were observed at the bottom of the control flasks that did not express car homologs. GC/MS analysis of the deposit dissolved in ethyl acetate (with 1% of acetic acid, v/v) revealed that the deposit was a fatty alcohol deposit.

60 C. Types of Fatty Alcohols Produced by Different CAR Homologs

Depending upon the CAR homolog expressed in *E. coli* strain C41 (DE3, ΔfadE), different mixtures of fatty alcohols were produced. Different compositions of fatty alcohols were 65 observed among the three CAR homologs evaluated (see Table 10). FadD9 produced more C<sub>12</sub> fatty alcohols relative to other fatty alcohols with carbon chain lengths greater than 12.

Both CarA and CarB produced a wider range in chain length of fatty alcohols than was observed when expressing FadD9.

Expected size of the  $\Delta$ fadD::Cm deletion was about 1200 bp (FIG. 4). The chloramphenicol resistance gene was elimi-

TABLE 10

Acyl-composition of fatty alcohols produced by recombinant <i>E. coli</i> strains							
Expressed with	Acyl-composition of fatty alcohols (%)						
TesA* and AlrAadp1	C10:0	C12	C14:1	C14:0	C16:1	C16:0	C18:1
CarA	trace	38	13	27	16	4	3
FadD9	trace	63	14	16	7	trace	trace
CarB	trace	32	11	41	12	trace	trace

\*the leaderless TesA, C12, including C12:0 and C12:1 fatty alcohol.

#### D. Quantification and Identification of Fatty Alcohols

Gas chromatography—mass spectrometry (GC/MS) was performed using an Agilent 5975B MSD system equipped with a 30 m×0.25 mm (0.10  $\mu$ m film) DB-5 column. The column temperature was 3 min isothermal at 100° C. The column was programmed to rise from 100° C. to 320° C. at a rate of 20° C./min. When the final temperature was reached, the column remained isothermal for 5 minutes at 320° C. The injection volume was 1  $\mu$ L. The carrier gas, helium, was released at 1.3 mL/min. The mass spectrometer was equipped with an electron impact ionization source. The ionization source temperature was set at 300° C.

Prior to quantification, various alcohols were identified using two methods. First, the GC retention time of each compound was compared to the retention time of a known standards, such as cetyl alcohol, dodecanol, tetradecanol, octadecanol, and cis-9-octadecenol. Second, identification of each compound was confirmed by matching the compound's mass spectrum to a standard's mass spectrum in the mass spectra library (e.g., C12:0, C12:1, C13:0, C14:0, C14:1, C15:0, C16:0, C16:1, C17:0, C18:0 and C18:1 alcohols).

#### Example 4

##### Production of Fatty Alcohol by Heterologous Expression of CAR Homologs in *E. coli* MG1655 (DE3, $\Delta$ fadD)

##### Construction of fadD Deletion Strain

The fadD gene of *E. coli* MG1655 was deleted using the lambda red system (Datsenko et al., 2000, Proc. Natl. Acad. Sci. USA, 97: 6640-6645) as follows:

The chloramphenicol acetyltransferase gene from pKD3 was amplified with the primers fad1 (5'-TAACCGCGTCT-GACGACTGACTAACGCTCAGGCTTAT-TGTCCACTTG TGTAGGCTGGAGCTGCTTCG-3') (SEQ ID NO:205), and fad2 (5'-CATTTGGGGTTCGAT-GACGACGAAACACGCATTAGAGGTGAAGAATTG CATATGAATATCCTCCTTAGTTCC-3') (SEQ ID NO:206).

This PCR product was electroporated into *E. coli* MG1655 (pKD46). The cells were plated on L-chloramphenicol (30  $\mu$ g/mL) (L-Cm) and grown overnight at 37° C. Individual colonies were picked on to another L-Cm plate and grown at 42° C. These colonies were then patched to L-Cm and L-carbenicillin (100 mg/mL) (L-Cb) plates and grown at 37° C. overnight. Colonies that were Cm<sup>R</sup> and Cb<sup>S</sup> were evaluated further by PCR to ensure the PCR product inserted at the correct site. PCR verification was performed on colony lysates of these bacteria using the primers fadF (5'-CGTC-CGTGGTAATCATTGG-3') (SEQ ID NO:207) and fadR (5'-TCGCAACCTTTCGTTGG-3') (SEQ ID NO:208).

nated using a FLP helper plasmid as described in Datsenko et al., Proc. Natl. Acad. Sci. USA 97:6640-6645 (2000). PCR verification of the deletion was performed with primers fadF and fadR (FIG. 4). The MG1655  $\Delta$ fadD strain was unable to grow on M9+oleate agar plates (oleate as carbon source). It was also unable to grow in M9+oleate liquid media. The growth defect was complemented by an *E. coli* fadD gene supplied in trans (in pCL1920-Ptrc).

##### Construction of MG1655(DE3, $\Delta$ fadD) Strain

To generate a T7-responsive strain, the  $\lambda$ DE3 Lysogenization Kit (Novagen) was utilized, which is designed for site-specific integration of  $\lambda$ DE3 prophage into an *E. coli* host chromosome, such that the lysogenized host can be used to express target genes cloned in T7 expression vectors.  $\lambda$ DE3 is a recombinant phage carrying the cloned gene for T7 RNA polymerase under lacUV5 control. Briefly, the host strain was cultured in LB supplemented with 0.2% maltose, 10 mM MgSO<sub>4</sub>, and antibiotics at 37° C. to an OD<sub>600</sub> of 0.5. Next, 10<sup>8</sup> pfu  $\lambda$ DE3, 10<sup>8</sup> pfu Helper Phage, and 10<sup>8</sup> pfu Selection Phage were incubated with 10  $\mu$ L host cells. The host/phage mixture was incubated at 37° C. for 20 min to allow phage to adsorb to host. Finally, the mixture was pipeted onto an LB plate supplemented with antibiotics. The mixture was spread evenly using plating beads, and the plates were inverted plates and incubated at 37° C. overnight.

$\lambda$ DE3 lysogen candidates were evaluated by their ability to support the growth of the T7 Tester Phage. T7 Tester Phage is a T7 phage deletion mutant that is completely defective unless active T7 RNA polymerase is provided by the host cell. The T7 Tester Phage makes very large plaques on authentic  $\lambda$ DE3 lysogens in the presence of IPTG, while much smaller plaques are observed in the absence of inducer. The relative size of the plaques in the absence of IPTG is an indication of the basal level expression of T7 RNA polymerase in the lysogen, and can vary widely between different host cell backgrounds.

The following procedure was used to determine the presence of DE3 lysogeny. First, candidate colonies were grown in LB supplemented with 0.2% maltose, 10 mM MgSO<sub>4</sub>, and antibiotics at 37° C. to an OD<sub>600</sub> of 0.5. An aliquot of T7 Tester Phage was then diluted in 1× Phage Dilution Buffer to a titer of 2×10<sup>3</sup> pfu/mL. In duplicate tubes, 100  $\mu$ L host cells were mixed with 100  $\mu$ L diluted phage. The host/phage mixture was incubated at room temperature for 10 min to allow phage to adsorb to host. Next, 3 mL of molten top agarose was added to each tube containing host and phage. The contents of one duplicate were plated onto an LB plate and the other duplicate onto an LB plate supplemented with 0.4 mM IPTG (isopropyl-β-thiogalactopyranoside) to evaluate induction of T7 RNA polymerase. Plates were allowed to sit undisturbed for 5 min until the top agarose hardened. The plates were then inverted at 30° C. overnight.

## Construction of MG1655(DE3, ΔfadD, yjgB::kan) Strain

The yjgB knockout strain, MG1655(DE3, ΔfadD, yjgB::kan), was constructed by using the following lambda red system (Datsenko et al., *Proc. Natl. Acad. Sci. USA* 97:6640-6645 (2000)):

The kanamycin resistant gene from pKD13 was amplified with the primers yjgBRn (5'-GCGCCTCAGATCAGCGCT-GCGAATGATTTCAAAAAATCGGCTTCAACAC-TGTAGGCTGGAGCTGCTCG-3') (SEQ ID NO:209), and yjgBFn (5'-CTGCCATGCTCTACACTTCCCAAACAA-CACCAGAGAAGGACCAAAAATG ATTCCGGG-GATCCGTCGACC-3') (SEQ ID NO:210). The PCR product was then electroporated into *E. coli* MG1655(DE3, ΔfadD)/pKD46. The cells were plated on kanamycin (50 µg/mL) (L-Kan) and grown overnight at 37° C. Individual colonies were picked on to another L-Kan plate and grown at 42° C. These colonies were then patched to L-Kan and carbenicillin (100 mg/mL) (L-Cb) plates and grown at 37° C. overnight. Colonies that were kan<sup>R</sup> and Cb<sup>S</sup> were evaluated further by PCR to ensure the PCR product was inserted at the correct site. PCR verification was performed on colony lysates of these bacteria using the primers BF (5'-gtgcggcataCGAACAAACA-3') (SEQ ID NO:211) and BR (5'-CCCCGC-CCTGCCATGCTCTACAC-3') (SEQ ID NO:212). The expected size of the yjgB::kan knockout was about 1450 bp.

## Evaluation of FadD on Fatty Alcohol Production Using MG1655(DE3, ΔfadD) Strain

In Example 3, a fadE deletion strain was used for fatty aldehyde and fatty alcohol production from 'TesA, CAR homologs, and endogenous alcohol dehydrogenase(s) in *E. coli*. To demonstrate that CAR homologs used fatty acids instead of acyl-CoA as a substrate, the gene encoding for acyl-CoA synthase in *E. coli* (fadD) was deleted so that the fatty acids produced were not activated with CoA. *E. coli* strain MG1655(DE3, ΔfadD) was transformed with pET-Duet-1-'tesA and pACYCDuet-1-carB. The transformants were evaluated for fatty alcohol production using the methods described herein. These transformants produced about 360 mg/L of fatty alcohols (dodecanol, dodecenol, tetradecanol, tetradeценол, cetyl, hexadecenol, and octadecenol).

## YjgB is an Alcohol Dehydrogenase

To confirm that YjgB was an alcohol dehydrogenase responsible for converting fatty aldehydes into their corresponding fatty alcohols, pETDuet-1-'tesA and pACYCDuet-1-fadD9 were co-transformed into either MG1655(DE3, ΔfadD) or MG1655(DE3, ΔfadD, yjgB::kan). At the same time, MG1655(DE3, ΔfadD, yjgB::kan) was transformed with both pETDuet-1-'tesA-yjgB and pACYCDuet-1-fadD9.

The *E. coli* transformants were grown in 3 mL of LB medium supplemented with carbenicillin (100 mg/L) and chloramphenicol (34 mg/L) at 37° C. After overnight growth, 15 µL of culture was transferred into 2 mL of fresh LB medium supplemented with carbenicillin and chloramphenicol. After 3.5 hrs of growth, 2 mL of culture was transferred into a 125 mL flask containing 20 mL of M9 medium with 2% glucose, carbenicillin, and chloramphenicol. When the OD<sub>600</sub> of the culture reached 0.9, 1 mM of IPTG was added to each flask. After 20 hrs of growth at 37° C., 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described herein.

The yjgB knockout strain resulted in significant accumulation of dodecanol and a lower fatty alcohol titer (FIG. 5). The expression of yjgB from plasmid pETDuet-1-'tesA-yjgB in the yjgB knockout strain effectively removed the accumulation of dodecanol (FIG. 5). The data shows that YjgB was

involved in converting dodecanal into dodecanol and that there may be other alcohol dehydrogenase(s) present in *E. coli* to convert other aldehydes into alcohols. Dodecanal accumulated in the yjgB knockout strain, but it was not observed in either the wild-type strain (MG1655(DE3, ΔfadD)) or the yjgB knockout strain with the yjgB expression plasmid. The arrows (in FIG. 5) indicate the GC trace of dodecanal (C12:0 aldehyde).

10

## Example 5

Production of Saturated Fatty Alcohols in *E. coli*

Fatty alcohols for commercial uses are saturated. However, 15 *E. coli* typically has a certain amount (about 20-25%) of unsaturated fatty acids in its membrane to maintain fluidity. An *E. coli* strain was engineered that was able to produce exclusively saturated fatty acids in a medium not supplemented with unsaturated fatty acid or cyclopropane-fatty acid 20 and was able to produce saturated fatty alcohols.

Two enzymes, a dehydratase/isomerase and a ketoacylsynthase I (KASI), encoded by fabA and fabB, respectively, are involved in unsaturated fatty acid biosynthesis. Usually, an *E. coli* strain lacking either FabA or FabB does not survive without supplementation of unsaturated fatty acids, such as oleate. To overcome this, the fabB gene was knocked out of an *E. coli* host strain, and the strain was able to grow without unsaturated fatty acid supplementation by genetically engineering the cells to express a recombinant desaturase gene 25 (AF037430, encoding DesA) from *Bacillus subtilis*. Although the first generation of the strain expressing desA required oleate for normal growth, subsequent plating of the strain on L Agar plates several times resulted in a strain that did not require oleate for growth.

## 35 Materials

*E. coli* JWC280 cells (described in Campbell et al., *Mol. Microbiol.* 47:793-805 (2003)) and *E. coli* GRT23 cells (described in Morgan-Kiss et al., *Arch. Microbiol.* 190:427-437 (2008)) were obtained from John Cronan.

## 40 Plasmid Construction

The desA gene (also referred to as Δ5 des) was amplified with primers delta5Fn and delta5Rn (listed in Table 11) from the genomic DNA of *Bacillus subtilis* str. 168 and digested with AvrII and EcoRI. The desA gene was then cloned into 45 pET-21(a), which had been linearized with AvrII-EcoRI, to produce pET-21a-Δ5. The desA gene was then removed as an NdeI-EcoRI fragment from pET-21a-Δ5 and inserted between the NdeI and EcoRI sites of OP180, a pACYC derived plasmid carrying a trc promoter. The resultant plasmid was named pACYC-Δ5.

A desA\_kan gene cassette was cloned between the AvrII-BamHI sites of CDFDuet-1. A kan gene cassette was produced by EcoRI and BamHI digestion of a PCR product that was amplified with primers kanF and kanR (see Table 11) 55 from pKD13 as the template (pKD13 was obtained from The Coli Genetic Stock Center, Yale University, and is described in Datsenko et al., *Proc. Natl. Acad. Sci. USA* 97:6640-6645 (2000)). The amplified desA gene (described above) was digested with AvrII and EcoRI. The AvrII-EcoRI fragment of the desA gene and the EcoRI-BamHI fragment of the kan gene were then inserted between the AvrII-BamHI sites of pCDFDuet-1 (from EMD Chemicals, Gibbstown, N.J.) to produce a plasmid that was named pCDFDuet-1-Δ5-kan.

A p84.17fabBA5kan plasmid was constructed to replace 60 fabB with the desA\_kan cassette by several subcloning steps. First, a DNA fragment (L-fabB) flanking the upstream region of fabB was amplified with primers fabBLF and fabBLR (see

Table 11), and a DNA fragment (R-fabB) flanking the downstream region of fabB was amplified with primers fabBRF and fabBRR (see Table 11) from *E. coli* MG1655 genomic DNA. Second, L-fabB was digested with XbaI and BglII, and R-fabB was digested with Nod and BglII. The digested L-fabB and R-fabB fragments were purified from agarose gel and were ligated with XbaI-NotI linearized pKOV. The resultant plasmid was designated pHZ1.186. Next, the desA\_kan gene cassette was removed from pCDFDuet-1-Δ5-kan as an AvrII-BamHI fragment and was inserted between the AvrII and BglII sites of pHZ1.186, resulting in the desA\_kan gene cassette being sandwiched by L-fabB and R-fabB. Finally, the L-fabB-desA\_kan-R-fabB fragment was amplified with fabBLF and fabBRR (see Table 11) from pHZ1.186 and cloned into the two PvuII sites of pMOD-4-MCS (Epicentre Biotechnologies, Madison, Wis.). The final plasmid was designated p84.17fabB.

DNA spanning from about 1 kb upstream to about 1 kb downstream of fabB::cm was amplified from the genome of GRT23 cells using the primers fabBup and fabBdown (see Table 11). The amplified DNA fragment was then digested with PvuII and inserted between the two PvuII sites of pMOD-4-MCS. The resulting plasmid was designated p84.15.

The genes encoding a thioesterase ('TesA) and a fatty acid reductase (CarB) were cloned as an operon, and the operon was placed under the trc promoter and pCL1920 vector. The final plasmid was named pCL-Ptrc-carB\_’tesA (the sequence is listed in FIG. 17 as SEQ ID NO:213).

TABLE 11

Primer sequences	
Primer ID	Sequence
delta5Fn	TTTT CCTAGG ATG ACT GAA CAA ACC A (SEQ ID NO: 214)
delta5Rn	TTTT GAATT C TTA TCA TTG TGA AAG CCAGAA (SEQ ID NO: 215)
kanF	TTTT GAATT C TGT AGG CTG GAG CTG CTTCG (SEQ ID NO: 216)
kanR	ATTCGG GGG ATC CGT CGA CC (SEQ ID NO: 217)
fabBLF	TTTT CTA GAA ATA GCG CCA GCG ACA (SEQ ID NO: 218)
fabBLR	TTTT AGA TCT TAG CCC TAG GCC AGT AAT CAC TGC ACG (SEQ ID NO: 219)
fabBRF	TTTT AGA TCT AGC TTC GGC TTC GGC G (SEQ ID NO: 220)
fabBRR	TTTT GCG GCC GCG CCC ATC CTT TGC TGG C (SEQ ID NO: 221)
fabBup	ACG ACA AAT GCG CCG C (SEQ ID NO: 222)
fabBdown	ATC CGC GCA ATA AAG C (SEQ ID NO: 223)

#### Strain Construction

An *E. coli* MG1655 ( $\Delta$ fadE $\Delta$ fhuAfabB::cm)/pACYC-Δ5 strain was constructed by transforming p84.15fabB into MG1655 ( $\Delta$ fadE $\Delta$ fhuAfabB::cm)/pACYC-Δ5. Plasmid p84.17fabB was transformed into MG1655 ( $\Delta$ fadE $\Delta$ fhuAfabB::cm)/pACYC-Δ5 to produce MG1655 ( $\Delta$ fadE $\Delta$ fhuAfabB::desA\_ kan)/pACYC-Δ5. After each transformation, the transformant mix was plated onto L agar plates supplemented with 1

mM IPTG and appropriate antibiotics (17 mg/L of chloramphenicol or 50 mg/L of kanamycin).

MG1655 ( $\Delta$ fadE $\Delta$ fhuAfabB::desA\_ kan)/pACYC-Δ5 grew normally in L Broth supplemented with oleate (potassium salt, 50 mg/L). Cells were plated onto L agar plates supplemented with 50 mg/L of oleate and incubated at 37°C. for 2 days. Colonies were then patched onto L Agar plates, supplemented with 50 mg/L of oleate and 100 mg/L of carbenicillin. One of the colonies, which lost resistance to carbenicillin but retained kanamycin resistance, was streaked onto an L agar plate supplemented with 50 mg/L of kanamycin, but no oleate. One of the colonies was selected from the plate and was designated ALC119A.

ALC119A with a Fatty Alcohol Pathway Produced Almost Exclusive Saturated Fatty Alcohol

Plasmid pCL-Ptrc-carB\_’tesA was transformed into the ALC119A strain. Three transformants of ALC119A/pCL-Ptrc-carB\_’tesA were grown in 3 mL of L broth with 100 mg/L of spectinomycin in a 37°C. shaker overnight. 15 μL of the overnight culture were transferred into 2 mL of fresh L broth with 100 mg/L of spectinomycin and 2 μL of 70% potassium oleate. The fresh inoculation was placed in a 37°C. shaker for about 3 hrs. The 2 mL culture was then transferred into 20 mL of V9 medium (Hu-9 medium without ferric chloride) in a 125 mL baffle flask. When the OD<sub>600</sub> of the culture reached about 0.9, 1 mM of IPTG was added to each flask. After 20 hrs of growth at 37°C., 20 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS as described in WO 2008/119082. Cetyl alcohol was used as a reference for quantification of fatty alcohol.

As shown in FIG. 13, the ALC119A/pCL-Ptrc-carB\_’tesA strain produced almost exclusively saturated fatty alcohols, including dodecanol, tetradecanol and hexadecanol.

#### Example 6

##### Production of Fatty Alcohols in the Cyanobacterium *Synechococcus* sp. PCC7002

This example describes the use of photoautotrophic bacteria to produce fatty alcohols from carbon dioxide (instead of glucose) using the carB\_’tesA-yahK pathway. First, a vector is constructed for homologous recombination into the *Synechococcus* sp. PCC7002 plasmid pAQ1 (genbank accession NC\_0050525) using 500 bp homologous regions corresponding to positions 3301-3800 and 3801-4300 of pAQ1. As a selectable marker, a spectinomycin resistance cassette containing the aminoglycoside 3' adenyltransferase, aad, promoter, gene and terminator (from plasmid pCL1920), is added between the homologous regions. For gene expression, the promoter and ribosome binding site of aminoglycoside phosphotransferase, aph (from plasmid pACYC177), is added followed by the unique cloning sites NdeI and EcoRI for insertion of a heterologous gene or operon. This complete integration cassette is constructed by gene synthesis and cloned into pUC19 for maintenance and delivery. The resulting plasmid, pLS9-7002, allows (i) cloning and expression of a foreign gene, and (ii) delivery and stable *in vivo* integration into *Synechococcus* sp. PCC7002 plasmid pAQ1.

The fatty alcohol pathway for expression in *Synechococcus* sp. PCC7002 is constructed as follows. The carB\_’tesA operon from pCL-Ptrc-carB\_’tesA (described in Example 4) is extended by adding yahK downstream of ‘tesA and then cloning into the NdeI and EcoRI sites of pLS9-7002 downstream of the aph promoter and ribosome binding site. The

resulting plasmid is transformed into *Synechococcus* sp. PCC7002 as described by Stevens et al. (*Proc. Natl. Acad. Sci. U.S.A.* 77:6052-6056 (1980)). Stable integrants are selected for on ATCC 1047 medium supplemented with 15 µg/mL spectinomycin. 1 L of ATCC 1047 medium contains 40 mg MgSO<sub>4</sub>×7H<sub>2</sub>O, 20 mg CaCl<sub>2</sub>×2 H<sub>2</sub>O, 750 mg NaNO<sub>3</sub>, 2 mg K<sub>2</sub>HPO<sub>4</sub>, 3.0 mg citric acid, 3.0 mg ferric ammonium citrate, 0.5 mg EDTA, 20 mg Na<sub>2</sub>CO<sub>3</sub>, 2.86 mg H<sub>3</sub>BO<sub>3</sub>, 1.81 mg MnCl<sub>2</sub>, 0.22 mg ZnSO<sub>4</sub>, 0.04 mg Na<sub>2</sub>MoO<sub>4</sub>, 0.08 mg CuSO<sub>4</sub>, 0.05 mg Co(NO<sub>3</sub>)<sub>2</sub>, 0.02 mg vitamin B12, 10 g agar, and 750 mL sea water. Spectinomycin resistant colonies are restreaked several times on ATCC medium 1047 with spectinomycin and tested for isogenic intergration of the carB-’tesA-yahK operon by PCR with primers pAQ1-U (atgtctgacaaggggtttgcaccc) (SEQ ID NO:224) and pAQ1-D (gcacatccttatccaattgccttag) (SEQ ID NO:225). Complete isogenic carB-’tesA-yahK integrants are then grown in 50 mL liquid ATCC 1047 medium with spectinomycin in 500 mL shake flasks with appropriate aeration and illumination at 30° C. for five to seven days. Culture aliquots are extracted at various time points with an equal volume of ethyl acetate and the extracts are analyzed for fatty alcohol production as described in Example 3. Fatty alcohols are produced.

#### Example 7

##### Production of Fatty Alcohols in the Cyanobacterium *Synechococcus elongatus* PCC7942

This example describes a second method of using photoautotrophic bacteria to produce fatty alcohols from carbon dioxide (instead of glucose) using the carB-’tesA-yahK pathway. First, a vector is constructed for homologous recombination into the *Synechococcus elongatus* PCC7942 genome (genbank accession CP\_000100) using 800 bp homologous regions corresponding to positions 2577844-2578659 and 2578660-2579467 of CP\_000100. This chromosomal location is known as neutral site one (NS1) (Mackey et al., *Meth. Mol. Biol.* 362:115-129 (2007)). As a selectable marker, a spectinomycin resistance cassette containing the aminoglycoside 3' adenyltransferase, aad, promoter, gene and terminator (from plasmid pCL1920), is added between the homologous regions. Additionally, the unique cloning sites NdeI and EcoRI are added for insertion of a heterologous gene or operon. This integration cassette is constructed by gene synthesis and cloned into pUC19 for maintenance and delivery. The resulting plasmid, pLS9-7942\_NS1, allows (i) cloning and expression of a foreign gene and (ii) delivery and stable *in vivo* integration into the *Synechococcus elongatus* PCC7942 genome.

The complete carB-’tesA-yahK operon (described in Example 6), including its ptrc promoter and ribosome binding site, is cloned into the NdeI or EcoRI site of pLS9-7942\_NS1. The resulting plasmid is transformed into *S. elongatus* PCC7942 as described by Mackey et al., *Meth. Mol. Biol.* 362:115-129 (2007). Stable integrants are selected for on BG-11 medium supplemented with 4 µg/mL spectinomycin. 1 L of BG-11 medium contains 75 mg MgSO<sub>4</sub>×7 H<sub>2</sub>O, 36 mg CaCl<sub>2</sub>×2 H<sub>2</sub>O, 1.5 g NaNO<sub>3</sub>, 40 mg K<sub>2</sub>HPO<sub>4</sub>, 6.0 mg citric acid, 6.0 mg ferric ammonium citrate, 1.0 mg EDTA, 20 mg Na<sub>2</sub>CO<sub>3</sub>, 2.86 mg H<sub>3</sub>BO<sub>3</sub>, 1.81 mg MnCl<sub>2</sub>, 0.22 mg ZnSO<sub>4</sub>, 0.04 mg Na<sub>2</sub>MoO<sub>4</sub>, 0.08 mg CuSO<sub>4</sub>, 0.05 mg Co(NO<sub>3</sub>)<sub>2</sub>, and 10 g agar. Spectinomycin resistant colonies are restreaked

several times on BG-11 medium with spectinomycin and tested for isogenic integration of the carB-’tesA-yahK operon by PCR with primers NS1-U (gatcaaacagggtgcagcagaactt) (SEQ ID NO:226) and NS1-D (attcttgacaagegatcgccgtcac) (SEQ ID NO:227). Complete isogenic carB-’tesA-yahK integrants are then grown in 50 mL liquid BG-11 medium with spectinomycin in 500 mL shake flasks with appropriate aeration and illumination at 30° C. up to seven days. Culture aliquots are extracted at various time points with an equal volume of ethyl acetate and the extracts are analyzed for fatty alcohol production as described in Example 3. Fatty alcohols are produced.

#### Example 8

##### Malonyl-CoA-Independent Production of Fatty Alcohols in *E. coli*

Certain protists such as *Euglena gracilis* are capable of malonyl-CoA independent fatty acid biosynthesis. The biosynthetic machinery for this pathway is located in the mitochondria and is thought to reverse the direction of β-oxidation by using acetyl-CoA as priming as well as elongating substrates to produce C<sub>8</sub> to C<sub>18</sub> fatty acids (Inui et al., *Eur. J. Biochem.* 142:121-126 (1984)). The enzymes involved are trans-2-enoyl-CoA reductases (TER), which catalyze the irreversible reduction of trans-2-enoyl-CoA to acyl-CoA and thereby drive the otherwise reversible pathway in the reductive direction (while the opposite is true for β-oxidation, where the irreversible acyl-CoA dehydrogenase, FadE, drives the reaction in the oxidative direction). One TER gene from *E. gracilis* as well as other eukaryotic and prokaryotic homologs are known (Hoffmeister et al., *J. Biol. Chem.* 280:4329-4338 (2005); Tucci et al., *FEBS Lett.* 581:1561-1566 (2007)). The only known TER enzyme from *E. gracilis* has been shown *in vitro* to reduce trans-2-butenoyl-CoA (C4) and trans-2-hexenoyl-CoA (C6) to the respective acyl-CoAs, (longer-chain trans-2-enoyl-CoAs have not been tested). Currently, very little is known about the other pathway enzymes in *E. gracilis*.

A pathway that creates a flux exclusively from acetyl-CoA precursors to acyl-CoA (as in *Euglena gracilis* mitochondria) can be engineered in *E. coli* using different sets of enzymes with the following four enzymatic activities: (i) non-decarboxylating, condensing thiolase, (ii) 3-ketoacyl-CoA reductase (or 3-hydroxyacyl-CoA dehydrogenase), (iii) 3-hydroxyacyl-CoA hydratase (or enoyl-CoA hydratase) and (iv) trans-2-enoyl-CoA reductase. All four enzymes can have sufficiently relaxed chain lengths specificity to allow synthesis of acyl-CoAs with longer chain length, e.g., C<sub>12</sub> or C<sub>14</sub>.

A plasmid encoding all four activities is constructed as follows. A synthetic operon of *E. coli* fadA (YP\_026272) (shown in FIG. 17 as SEQ ID NO:229) (non-decarboxylating thiolase) and fadB (NP\_418288) (shown in FIG. 17 as SEQ ID NO:231) (3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase) and *E. gracilis* ter (Q5EU90) (shown in FIG. 17 as SEQ ID NO:228) (trans-2-enoyl-CoA reductase, codon optimized without its 5' sequence encoding a transit peptide) is constructed and cloned downstream of a ptrc promoter into a pACYC plasmid with a carbenicillin or chloramphenicol resistance gene. Alternatively, instead of the *E. coli* fadA and fadB genes, the *E. coli* fadI (NP\_416844) (shown in

FIG. 17 as SEQ ID NO:223) and fadJ (NP\_416843) (shown in FIG. 17 as SEQ ID NO:235) genes (or the corresponding orthologs from other organisms) are used. As an alternative to the *E. gracilis* ter gene, the corresponding orthologs from other organisms or the *E. coli* fabI (NP\_415804) (shown in FIG. 17 as SEQ ID NO:237) gene are used. Although FabI normally reduces trans-2-enoyl-ACPs, it is also active with trans-2-enoyl-CoAs (Bergler et al., *J. Biol. Chem.* 269:5493-5496 (1993)).

The pACYC-ptrc\_fadAB-ter plasmid or the pACYC-ptrc\_fadAB-fabI plasmid is cotransformed with the pCL-ptrc\_carB-'tesA plasmid (described in Example 4) into an *E. coli* ΔfadE strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3. The two different strains produce fatty alcohols with different chain length distribution.

As these strains express 'TesA, a portion of the fatty alcohols produced are derived from malonyl-CoA dependent acyl-ACP precursors. 'TesA efficiently hydrolyzes acyl-ACPs when overexpressed in *E. coli*, although it has higher specific activity for acyl-CoAs as compared to acyl-ACPs. To increase the proportion or exclusively produce fatty alcohols derived from the malonyl-CoA independent pathway, alternative thioesterases that have lower hydrolytic activity towards acyl-ACPs are used instead of 'TesA. One example is *E. coli* TesB (NP\_414986), which prefers acyl-CoAs over acyl-ACPs (Spencer et al., *J. Biol. Chem.* 253:5922-5926 (1978)) and when overexpressed in *E. coli* does not hydrolyze acyl ACPS (Zheng et al., *App. Environ. Microbiol.* 70:3807-3813 (2004)). In alternative methods, orthologs of TesA and TesB or thioesterases from other protein families that hydrolyze acyl-CoAs with high efficiency while hydrolyzing acyl-ACPs with low efficiency are used.

In one method, a pCL-ptrc\_carB-'tesB plasmid is constructed as described in Example 4 by replacing the 'tesA gene with the tesB gene (NP\_414986) (shown in FIG. 17 as SEQ ID NO:239). The plasmid is cotransformed with the pACYC-ptrc\_fadAB-ter plasmid or the pACYC-ptrc\_fadAB-fabI plasmid into an *E. coli* ΔfadE strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3.

In another method, the pCL-ptrc\_carB-'tesA plasmid is replaced with a pCL-ptrc\_acr1 plasmid, which expresses the acyl-CoA reductase Acr1 from *Acinetobacter baylyi* ADP1 (YP\_047869) (shown in FIG. 17 as SEQ ID NO:241). This reductase specifically reduces acyl-CoAs but not acyl-ACPs to the corresponding fatty alcohols (Reiser et al., *J. Bacteriol.* 179:2969-2975 (1997)). The plasmid is cotransformed with the pACYC-ptrc\_fadAB-ter plasmid or the pACYC-ptrc\_fadAB-fabI plasmid into an *E. coli* ΔfadE strain. These strains are cultured, extracted and analyzed for fatty alcohol production as described in Example 3. The strains produce fatty alcohols independent of malonyl-CoA.

#### Example 9

##### Identification of Iron as an Inhibitor of Fatty Alcohol Production

Hu9 medium is a known fermentation medium, which contains 6 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/L NaCl, 1 g/L

NH<sub>4</sub>Cl, 0.1 mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 15 g/L agar, 10 mM glucose, 50 mg/L Uracil, and trace minerals containing 100 μM FeCl<sub>3</sub>, 500 μM ZnCl<sub>2</sub>, 200 μM Na<sub>2</sub>Mo<sub>4</sub>, 200 μM CuSO<sub>4</sub>, 200 μM H<sub>3</sub>BO<sub>3</sub>. However, it was observed that the production of fatty alcohols was completely reduced when recombinant *E. coli* strains, otherwise capable of producing fatty alcohols, were grown in Hu9 medium. As described in detail below, the inability of *E. coli* strains to produce fatty alcohols in various incomplete Hu9 media was measured, and it was found that the recombinant bacteria were incapable of producing fatty alcohols when iron was present in the medium. However, the addition of iron did not inhibit the growth of the bacteria.

In order to identify the component(s) involved in the inhibition of fatty alcohol production, different versions of incomplete Hu9 medium were made, some of which lacked a dispensable ingredient, and then the production of fatty alcohol was evaluated.

In the first step the following media were made: complete Hu9 medium, incomplete Hu9 medium lacking uracil, and incomplete Hu9 medium lacking trace elements. K6 cells (a recombinant bacterial strain C41 (DE3, ΔfadE) carrying pACYCDuet-1-carB, encoding the CAR homolog carb and pETDuet-1-'tesA) were cultured in 2 mL of LB containing appropriate antibiotics. After reaching an OD of 1.0, the 2 mL cultures were scaled up in 125 mL shake flasks (containing one of the Hu9 media described above) to a volume 22 mL. The cultures were induced by adding IPTG to a final concentration of 1 mM. After growing them for 20 hrs at 37° C., 22 mL of ethyl acetate (with 1% of acetic acid, v/v) was added to each flask to extract the fatty alcohols produced during the fermentation. The crude ethyl acetate extract was directly analyzed with GC/MS and the total fatty alcohol titers were quantified.

As depicted in FIG. 14A, the fatty alcohol production was inhibited to a great extent by the addition of trace elements as compared to the addition of uracil to the incomplete Hu9 medium. This indicated that the inhibitory component(s) was a part of trace mineral solution.

In order to find out which trace element was responsible for the fatty alcohol production inhibition, the following Hu9 media were made: complete Hu9 medium; Hu9 lacking FeCl<sub>3</sub>; Hu9 lacking ZnCl<sub>2</sub>; Hu9 lacking Na<sub>2</sub>Mo<sub>4</sub>; Hu9 lacking CuSO<sub>4</sub>; and Hu9 lacking H<sub>3</sub>BO<sub>3</sub>. The fatty alcohol production of K<sub>6</sub> cells grown in these different Hu9 media was evaluated using the method described above.

As shown in FIG. 14B, fatty alcohol production was inhibited mainly by the addition of iron to the medium. Thus, by eliminating or reducing the presence of iron (e.g., ferric citrate, ferric chloride, or ferrous sulfate) in the culture medium, fatty alcohols can be produced.

#### 60 Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

## SEQUENCE LISTING

The patent contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US09068201B2>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

The invention claimed is:

**1. A bacterial cell culture comprising:**

(a) a bacterial cell genetically engineered to express an exogenous polypeptide having the amino acid sequence of SEQ ID NO:22, and  
 (b) a glycerol-containing carbon source,  
 wherein a fatty alcohol composition comprising C12 and C14 fatty alcohols is produced by the genetically engineered bacterial cell.

**2. The cell culture of claim 1, wherein the bacterial cell is further engineered to express an exogenous polypeptide having thioesterase (EC 3.1.2.14 or EC 3.1.1.5) activity effective to release a fatty acid from acyl-ACP.**

**3. The cell culture of claim 1, wherein acyl-CoA dehydrogenase (fadE; EC 1.3.99.3, 1.3.99.) expression or activity is attenuated in the bacterial cell as compared to a wild-type bacterial cell.**

**4. The cell culture of claim 1, wherein acyl-CoA dehydrogenase (YdiO; EC 1.3.99.) expression or activity is deleted or reduced in the bacterial cell as compared to a wild-type bacterial cell.**

**5. The cell culture of claim 1, wherein 3-ketoacyl-CoA thiolase (fadA; EC 2.3.1.16) expression or activity is deleted or reduced in the bacterial cell as compared to a wild-type bacterial cell.**

**6. The cell culture of claim 1, wherein fadB (EC 4.2.1.17, EC 5.1.2.3, EC 5.3.3.8, EC 1.1.1.35) expression or activity is deleted or reduced in the bacterial cell as compared to a wild-type bacterial cell.**

**7. The cell culture of claim 1, wherein beta-ketoacyl-CoA thiolase (fadI; EC 2.3.1.16) expression or activity is deleted or reduced in the bacterial cell as compared to a wild-type bacterial cell.**

**8. The cell culture of claim 1, wherein fadJ (EC 5.1.2.3, EC 4.2.1.17, EC 1.1.1.35) expression or activity is deleted or reduced in the bacterial cell as compared to a wild-type bacterial cell.**

**9. A bacterial cell genetically engineered to express**  
 (a) an exogenous polypeptide having the amino acid sequence of SEQ ID NO:22,  
 (b) a polypeptide having thioesterase (EC 3.1.2.14 or EC 3.1.1.5) activity, and  
 (c) a polypeptide having attenuated acyl-CoA dehydrogenase (fadE; EC 1.3.99.3, EC 1.3.99.) activity,  
 wherein the genetically engineered bacterial cell produces a fatty alcohol composition comprising C12 and C14

fatty alcohols when cultured in the presence of a glycerol-containing carbon source.

**10. A method of making a fatty alcohol composition, comprising:**

a. culturing a bacterial cell genetically engineered to express an exogenous polypeptide having the amino acid sequence of SEQ ID NO:22 in a culture medium comprising glycerol;  
 b. fermenting the genetically engineered bacterial cell of part a in a fermentation media, wherein a fatty alcohol composition comprising C12 and C14 fatty alcohols is produced by the genetically engineered bacterial cell; and  
 c. optionally separating the fatty alcohol composition from the fermentation media.

**11. The method of claim 10, wherein the genetically engineered bacterial cell is further engineered to express an exogenous polynucleotide sequence encoding a polypeptide having thioesterase (EC 3.1.2.14 or EC 3.1.1.5) activity effective to release a fatty acid from acyl-ACP.**

**12. The method of claim 10, wherein the culture medium further comprises a fatty acid.**

**13. The method of claim 10, wherein, acyl-CoA dehydrogenase (fadE; EC 1.3.99.3, 1.3.99.) expression or activity is attenuated in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

**14. The method of claim 10, wherein acyl-CoA dehydrogenase (YdiO; EC 1.3.99.) expression or activity is deleted or reduced in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

**15. The method of claim 10, wherein 3-ketoacyl-CoA thiolase (fadA; EC 2.3.1.16) expression or activity is deleted or reduced in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

**16. The method of claim 10, wherein fadB (EC 4.2.1.17, 5.1.2.3, 5.3.3.8, 1.1.1.35) expression or activity is deleted or reduced in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

**17. The method of claim 10, wherein beta-ketoacyl-CoA thiolase (fadI; EC 2.3.1.16) expression or activity is deleted or reduced in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

**18. The method of claim 10, wherein fadJ (EC 5.1.2.3, 4.2.1.17, 1.1.1.35) expression or activity is deleted or reduced in the genetically engineered bacterial cell as compared to a wild-type bacterial cell.**

\* \* \* \* \*